SCOTTISH JOURNAL OF POLITICAL ECONOMY

VOLUME 43 FEBRUARY 1996 NO. 1

Articles

S. C. Littlechild	PRIVATISATION, COMPETITION AND REGULATION IN THE SCOTTISH ELECTRICITY INDUSTRY	1
David Blake	FINANCIAL INTERMEDIATION AND FINANCIAL INNOVATION IN A CHARACTERISTICS FRAMEWORK	16
Paul Mizen and John R. Presley	THE REAL BALANCE EFFECT AND STABILITY ANALYSIS IN CLASSICAL MONETARY THEORY	32
Deborah Redman	SIR JAMES STEUART'S STATESMAN REVISITED IN LIGHT OF THE CONTINENTAL INFLUENCE	48
Patricia Fraser	UK EXCESS SHARE RETURNS: FIRM SIZE AND VOLATILITY	71
Paolo M. Panteghini	SUNK COSTS AND PROFIT TAXATION: A SOURCE OF TAX NON-NEUTRALITY	85
Donald A. R. George	THE PRICE-QUALITY RELATIONSHIP UNDER MONOPOLY AND COMPETITION	99
Colin Wren	GROSS EXPENDITURE ON UK INDUSTRIAL ASSISTANCE: A RESEARCH NOTE	113
Short Reviews		127

THE SCOTTISH ECONOMIC SOCIETY

President:	Professor Sir Laurence Hunter
Vice-Presidents:	Sir Fraser Noble Professor M. Gaskin Sir Alex Cairncross Professor D. Flint Dr T. L. Johnston Professor T. Wilson
Editors:	Professor L. C. Hunter and Professor A. Muscatelli
Review Editor:	Professor A. Dow
Secretary:	F. G. Hay
Treasurer:	J. Houston

Editorial Board

Brian K. Ashcroft	Neil M. Kay	Charlotte Lythe
John Foster	Alex G. Kemp	Gavin C. Reid
Peter H. Grinyer	Brian J. Loasby	Andrew S. Skinner

Scottish Journal of Political Economy is published five times a year in February, May, August, September and November by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF or 238 Main Street, Cambridge, MA 02142, USA.

Information for Subscribers The subscription rates for 1996 are £78.00 UK/Europe, £90.00 ROW, $144.00 North America*. The single issue price is £18.00 UK/Europe, £21.00 ROW, $33.00 North America*. New orders and sample copy requests should be sent to the Journals Marketing Manager at the Publisher; renewals, claims and other subscription enquiries to the Journals Subscriptions Department, Marston Book Services, PO Box 270, Abingdon, Oxon OX4 4YW. Cheques should be made payable to Blackwell Publishers Ltd. All subscriptions are supplied on a Calendar year basis (January to December).
*Canadian customers/residents please add 7% for GST.

The journal is distributed free of charge to members of the Scottish Economic Society. Applications for membership of the society, normally available only to individuals, should be sent to the Treasurer, Mr J. Houston, Department of Economics, Glasgow Caledonian University, 70 Cowcaddens Road, Glasgow G4 0BA. The annual subscription to the Society for members is £22.00 (US$35.00, North America). Student membership is £13.00 (US$22.00 North America) and family membership is £7.00 (US$13.00 North America). Correspondence for the Society on other matters should be addressed to the Secretary, Mr F. G. Hay, Department of Political Economy, University of Glasgow G12 8QQ.

The journal is available on microfilm (16mm or 35mm) or 105mm microfiche from the Serials Acquisitions Department, University Microfilms Inc., 300 North Zeeb Road, Ann Arbor, MI 48106, USA.

Back Issues Single issues from the current and previous 3 volumes are available from Marston Book Services at the address above at the current single issue rate. Earlier issues may be obtained from Dawsons Back Issues, Cannon House, Folkestone, Kent CT19 5EE.

US Mailing Second class postage paid at Rahway, New Jersey. Postmaster: send address corrections to *Scottish Journal of Political Economy*, c/o Mercury Air Freight International Ltd. Inc., 2323 E-F Randolph Avenue, Avenel, NJ 07001, USA. (US mailing agent)

Abstracting and Indexing This journal is indexed or abstracted in Journal of Economic Literature and Work Related Abstracts.

© 1996 Scottish Economic Society. All rights reserved. Apart from fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may not be reproduced, stored or transmitted in any form or by any means without the prior permission in writing of the publisher, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency or the Copyright Clearance Center.

Blackwell Publishers

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

PRIVATISATION, COMPETITION AND REGULATION IN THE SCOTTISH ELECTRICITY INDUSTRY†

*S. C. Littlechild**

ABSTRACT

This paper analyses the experience of the privatised electricity industry in Scotland. The retention of vertical integration in Scotland has led to a number of different features in the regulation of the industry in Scotland and England and Wales. It has also led to a different pattern of development of competition. Although regulation and price competition appear to have operated to the benefit of consumers thus far, the Scottish system is as yet untested in its ability to cope with significant new entry.

I INTRODUCTION

It is a great honour to be asked to give the President's lecture at your conference this year. It is also a particular pleasure to be here in Aberdeen during the University's celebration of its quincentenary. What I would like to do this afternoon is discuss something of rather more recent vintage than the University, namely the structure and regulation of the privatised electricity industry. I would like to take the opportunity of addressing a Scottish audience to focus particularly on developments in the industry in Scotland.

II BACKGROUND

Before privatisation the electricity industry in Scotland was structured differently from that in England and Wales. Those differences have in large part, been carried through into the industry since privatisation. They have had implications for the regulatory regime and also for the development of competition in Scotland.

I would draw attention to five main features. First, in England and Wales all generation and transmission was concentrated in one organisation, the Central Electricity Generating Board, while distribution and supply was split between

† President's Lecture, Scottish Economic Society Annual Conference, Conference Centre, University of Aberdeen, 13 April 1995.

* Director General of Electricity Supply.

12 Area Boards. In contrast, the industry in Scotland was characterised by two vertically integrated boards, each covering a different part of the country: the North of Scotland Hydro-Electricity Board and the South of Scotland Electricity Board.

Second, it was nonetheless held that planning and operating the system on a Scotland-wide basis could assist in minimising the running costs of the system. For example, a Joint Generating Agreement between the two companies ensured a single merit order and the sharing of reserve.

Third, planning the system on a Scotland-wide basis led to the position where neither company had a balanced portfolio of generation plant. The SSEB owned all of the coal and nuclear plant. The NSHEB owned the greater part of the conventional hydro generation, all the pumped storage plant, and the dual-fired station at Peterhead.

Fourth, it was decided to build plant in Scotland that reaped the benefits of economies in plant construction. This meant that Scotland had relatively few power stations, but each station was relatively large and could supply a significant proportion of Scottish demand.

Fifth, investment decisions had endowed Scotland with a substantial surplus of capacity over demand.

At privatisation the Government took the decision to leave the existing structure in Scotland largely intact. This was in marked contrast to the reorganisation of the industry in England and Wales. The CEGB, which had been responsible for all generation and transmission, was split into three generating companies (National Power, PowerGen and Nuclear Electric) and a separate transmission company (National Grid Company). The latter was made responsible for operating the high voltage national transmission network throughout England and Wales.

Transmission was not separated out in Scotland: ScottishPower and the Hydro-Electric each separately retain ownership and responsibility for operating the transmission systems in their own area, and new separate generating companies were not formed, with the exception of the creation of Scottish Nuclear to own and operate the SSEB's nuclear plant. This followed the decision to leave nuclear plant out of the privatisation in England and Wales. However, unlike England and Wales there was no provision for Scottish Nuclear to compete by selling directly to the market. Instead, it entered into a contract for all its output with ScottishPower and Hydro-Electric, for the next 15 years.

In leaving the industry structure largely intact, Government indicated that it was conscious of the relatively small size of the Scottish system, only around one-eighth of the size of that in England and Wales. It also considered that vertical integration had been a successful industry structure in Scotland, and that it had particular beneficial features in serving sparsely populated areas. The Government hoped that direct competition between the two companies, for example for large industrial customers, and indirect competition by comparison, would allow the Scottish industry to reap the efficiency benefits of private ownership and competition without the need for fundamental restructuring. It

© Scottish Economic Society 1996

was also recognised, however, that cooperative arrangements had the effect of diluting the incentives on the management of each company to control costs.

Restructuring and privatisation of the industry therefore had the twin objectives of sharpening the commercial incentives on management whilst ensuring that where there were clear benefits to cooperation, those benefits could be harnessed within the privatised structure.[1]

To provide each of the companies with a balanced mix ofgenerating plant a complex series of contracts, collectively known as the restructuring contracts, was put in place (Figure 1). Thus Hydro-Electric has access to a share of ScottishPower's coal-fired plant; and ScottishPower has access to a share of the Peterhead station and of Hydro-Electric's conventional hydro plant. In addition, the Cruachan pumped-storage plant was transferred from Hydro-Electric to ScottishPower. These contracts reduce the reliance of each company on any one fuel, or on the performance of individual stations. They also have the effect of giving the companies very similar costs at the margin. This may, in turn, influence the degree of competition that is likely or possible between the two companies.

In England and Wales, the separation of generation from supply led to the creation of a trading system for generation—The Pool. No such arrangement was put in place in Scotland, and different arrangements were to be developed, as discussed shortly.

The distinctive features of the industry structure have been recognised in the form of regulation for Scotland. The terms of the Electricity Act, the legislation that brought about the restructuring of the industry, and which set the framework for its regulation, apply equally to Scotland as to England and Wales. However, the licences under which the companies operate are different

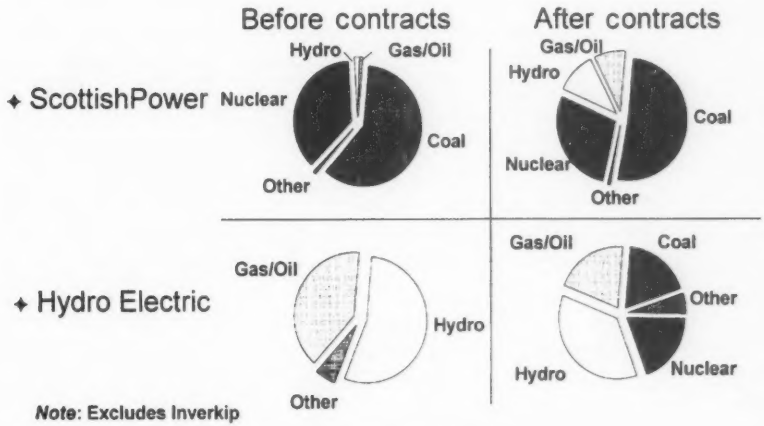


Figure 1. Restructuring contracts.

[1] 'Privatisation of the Scottish Electricity Industry' (Cm 327, HMSO, 1988).

© Scottish Economic Society 1996

in Scotland from those in England and Wales. ScottishPower and Hydro-Electric have composite licences which encompass the activities of transmission, generation and public electricity supply. There are also certain features of those licences that differ from their equivalents in England and Wales. I will give three illustrations.

First, public electricity supply licences contain different forms of regulation of generation prices. In England and Wales there are no price controls in the licences of the generation companies. Subject to an obligation to purchase economically, the RECs as suppliers are able to pass-through to customers the cost of generation purchases. This reflects the fact that they are purchasing from a relatively competitive generation market and are separately restricted in the amount of generation capacity they themselves can own. In Scotland, by contrast, the companies as suppliers are predominantly purchasing from their own generation sources. They too are subject to an economic purchasing condition, but the licences also impose an explicit limit on the price they can charge in respect of generation.

A second difference is in the nature of the arrangements that allow for competitive suppliers to operate in the market. In England and Wales the Pool is a market that provides a mechanism by which all suppliers are able to gain access to generation, and a basis for settling accounts. In Scotland no such market presently exists. The licences therefore require the two companies to make arrangements that allow second-tier suppliers to operate in the Scottish market. Those arrangements are subject to my approval and can have an important influence on the extent and nature of supply competition in Scotland.

The present arrangements are relatively simple. They allow a generator to contract directly with customers—making use of the transmission and distribution system on published, non-discriminatory terms. However, where a generator supplies a customer directly in this way, there is no means of ensuring that the generator's output will exactly match his customer's demand. So some mechanism is therefore needed to account for the periods of over and under-supply by the generator. Such differences between the profile of the generator's output and the profile of his customers' demand, are settled by reference to the England and Wales Pool price. Suppliers who are not generators are allowed to purchase their requirements from any generator but are also entitled to purchase from the local public suppliers on the basis of Pool price. In practice suppliers in Scotland have purchased all their generation requirements from Scottish-Power and Hydro-Electric as their local supplier, typically at Pool price. This is a point to which I return below.

A third difference pertains to transmission. In England and Wales, NGC has only limited involvement in generation, through its ownership of pumped-storage plant, and no involvement in supply. In Scotland, by contrast, ScottishPower and Hydro-Electric are involved in the competitive activities of supply and generation, with access to potentially commercially valuable information in their role as operators of the grid. So a 'Chinese Wall' condition applies to ScottishPower's and Hydro-Electric's transmission licence. This requires each company to make arrangements, again subject to my approval, to

© Scottish Economic Society 1996

prevent the use of commercially sensitive information about generators and suppliers to further the licensee's own interests in these competitive businesses.

It is apparent from these examples that there was a recognition at privatisation that continued vertical integration might have anti-competitive effects unless explicit regulatory mechanisms were put in place to help enforce a degree of separation between monopoly and competitive activities.

These important differences notwithstanding, the Scottish market was expected to develop in a similar manner to that in England and Wales with the introduction of competition in generation, a programme for the introduction of competition in supply, and the continuation of price controls on monopoly activities. We now turn to an examination of how these different arrangements have developed in practice.

III Competition in Generation

In England and Wales, a key feature of the first years of privatisation has been changing market shares between existing generators and entry of new generating companies. Smaller generators have expanded output. Older, more costly plant has closed and new, lower cost, more efficient gas plant has come on to the system.

Figure 2 shows the changing market shares of output of the major generators in England and Wales. By improving availability and performance Nuclear Electric increased market share by nearly 50% between 1989/90 and 1994/95, and its share is likely to increase further with the opening of Sizewell 'B'. The two Scottish companies and Electricité de France (EDF) have also increased their output through the Scottish and French interconnectors, again by about 50%. And new entrants' gas plants now provide about 8% of the market, with more new plant under construction. In parallel with this, the two major fossil

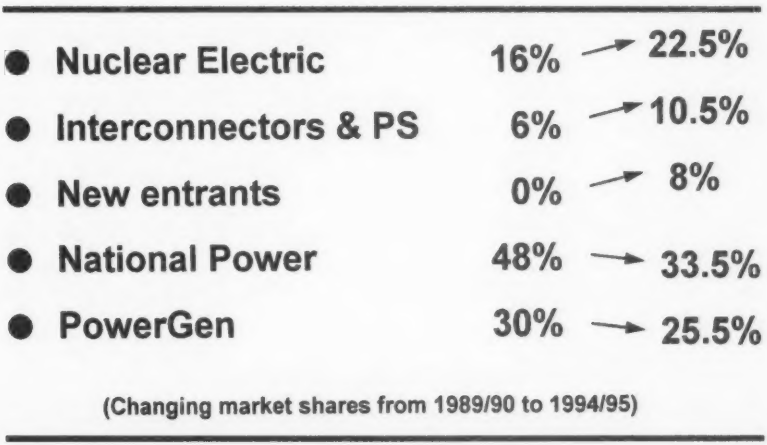

Nuclear Electric 16% → 22.5%
Interconnectors & PS 6% → 10.5%
New entrants 0% → 8%
National Power 48% → 33.5%
PowerGen 30% → 25.5%
(Changing market shares from 1989/90 to 1994/95)

Figure 2. Competition in generation England and Wales.

© Scottish Economic Society 1996

generators, National Power and PowerGen, have conceded a combined reduction in market share of nearly 20 percentage points. In Scotland, the first notable feature is the substantial overall increase in output of Scottish generation—up 28% between 1989/90 and 1994/5. Whilst Scottish demand has grown at only a modest rate, exports to England and Wales have increased substantially. They now account for around 20% of total Scottish generation compared to about typically 5% before privatisation. This is reflected in the increase in the market share of the interconnectors shown in Figure 2.

In other respects, the generation system in Scotland is much like it was five years ago. SNL's nuclear station Hunterston 'A' closed at privatisation and the AEA's prototype fast-breeder reactor has since closed. SNL's nuclear (AGR) station at Torness began production a little before privatisation. Torness and SNL's other station at Hunterston 'B' have increased output significantly over the past five years, offsetting much of the loss of output that followed the closure of Hunterston 'A'. Their increase in output of 35% over the past five years compares with Nuclear Electric's increase in output of around 50%. However, in an expanding market, SNL's market share has fallen from 54% to 43%. Over this period, the introduction of sour gas from the Miller field has permitted substantial increases in output from the Peterhead station, which is shared by ScottishPower and Hydro-Electric, and the market shares of these two companies have each increased by about 5 percentage points. There has thus been only one significant change in the generation capacity of existing companies. And there has been no significant new entry (Figure 3). Changes in market share have been almost entirely as a result of increased or reduced output from existing plant.

IV Competition in Supply

Now let us turn to competition in supply—that is, the ability of generators and other suppliers to compete to supply customers directly using the transmission and distribution lines of the public electricity suppliers. There has been competition in supply since 1990 for larger customers, those with an electricity demand greater than 1 MW. Throughout Great Britain there are some 5,000 customers in this category, just under 500 of those are in Scotland. Since April

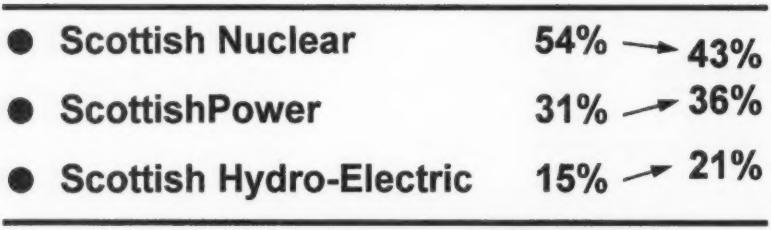


Figure 3. Competition in generation Scotland.

© Scottish Economic Society 1996

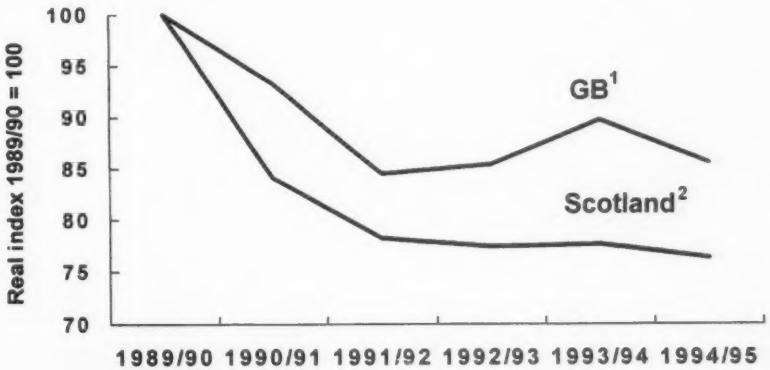
last year customers with maximum demand above 100 kW have also been free to choose their electricity supplier. This has increased the number of customers in the competitive market by about ten-fold (to about 50,000) and has taken total volume in the competitive market to around 50% of total demand. In 1998 the market will be fully open to competition and all customers, right down to the domestic level, will be free to choose their electricity supplier.

This is perhaps the most distinctive feature of the new arrangements in the electricity supply industry in Great Britain. In 1990 no other country (except Chile to a limited extent) had made an attempt to introduce competition at the retail level. Even now, relatively few countries have adopted this principle, although support for it is growing.

Effect on price

Let us first examine the effect of competition on price (Figure 4). The experience of customers who have had a choice of supplier, both in Scotland and in England and Wales, has generally been of appreciable reductions in price. Information collected by the Government, from a survey of manufacturing sites throughout Great Britain, provides information on prices paid by groups of customers in Great Britain as a whole. These show that most large customers (an average of the sites classed as Medium and Large) enjoyed price reductions in the two years after privatisation of around 15%. There was some slight increase in prices since then, but in real terms the level of prices to most large users is still around 15% below that of five years ago.

Some very large customers in England and Wales did experience price increases at privatisation, principally as a result of the ending of favoured terms they had previously enjoyed. Nevertheless, average prices to those customers as a group (the Extra Large sites) are now roughly back to their pre-privatisation level in real terms.



[1] Source, Energy Trends.
[2] Source, price control data.

Figure 4. Prices to non-franchise customers.

© Scottish Economic Society 1996

Information we collect from price control returns made by the Scottish companies enables us to calculate the prices paid by customers in the market over 1 MW in Scotland. This information shows that in the two years after privatisation prices to such customers fell on average by over 20% in real terms. Prices have fallen a little further since then and are now nearly a quarter lower in real terms than they were in 1989/90. These figures need to be treated with some caution as they do not directly address the possibility of different price levels before privatisation in England and Wales and Scotland and the data sources may not be directly comparable. Nonetheless, they suggest that the degree of price competition in the competitive market in Scotland was no less than that in England and Wales, and may have been greater.

This conclusion is consistent with OFFER's experience in discussing the market with large customers. We have had relatively fewer complaints about the level of prices in Scotland, and customers with knowledge of both the English and Scottish markets have tended to comment that prices in Scotland compare favourably with England and Wales.

Customers in the above 100 kW market have been in the competitive market for just a year. Last April, they reported price reductions of around 10% on average compared to 1993/94. In Scotland, we are again able to calculate the extent of reduction in price from price control data submitted by the Scottish companies. This confirms a fall of around 10% in the financial year just completed. Some further reductions have been reported this year in both parts of Great Britain, partly but not entirely by those customers who did not negotiate such significant reductions last year.

Effect on choice of supplier

I would now like to turn to consider how competition has manifested itself in terms of changing market shares of suppliers. Here the position is very different between Scotland and England and Wales. Figure 5 shows the proportion of

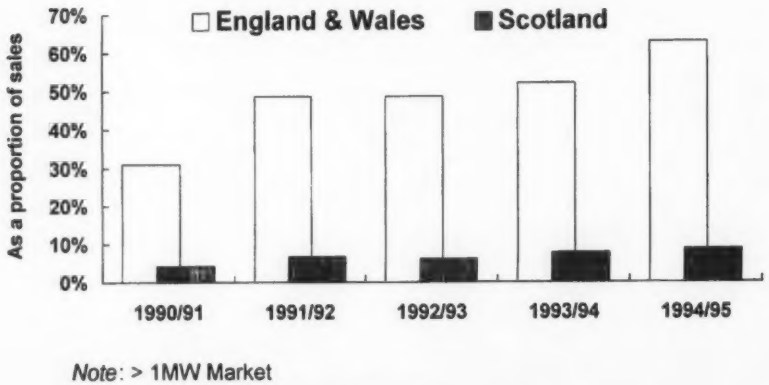


Figure 5. Market share 2nd tier supply.

© Scottish Economic Society 1996

customers' demand in the over 1 MW market in England and Wales that is supplied other than by the local supplier for each year since privatisation. Immediately upon the introduction of competition a significant proportion of customers in England and Wales switched to an alternative—second-tier—supplier. The extent of that switch was constrained by limits at the time on direct sales by National Power and PowerGen. Even so, in 1990/91 27% of customers, representing 31% of demand in the competitive market, left their local supplier. In each year since 1990 the volume of demand supplied second-tier has increased. Last year, nearly two thirds of over 1 MW demand was met by suppliers other than the local supplier. The three main generators have established a large and slightly increasing market share, which reached about 40% last year.

The position in Scotland has been quite different. The proportion of demand met by second-tier suppliers has remained roughly constant at around 6% since privatisation. Moreover, practically all of the second-tier market share in Scotland has been accounted for by the other Scottish company securing customers from its neighbour. All except one of the regional electricity companies from England and Wales now have a licence to supply in Scotland as do National Power, PowerGen and Nuclear Electric. Not all appear to be particularly active in seeking customers in Scotland but a number have been. However, the success of England and Wales suppliers in securing customers has been very limited.

It has therefore been a matter of some interest why the pattern of competition should have been quite so different. One possibility is that the profit margin on supply alone is slim so that, for the REC incumbents in England and Wales, supply has limited value. However, for the incumbents in Scotland their profit margin in generation—indeed their share of the Scottish generation market as a whole—is also at stake. In other words, by virtue of their vertical integration ScottishPower and Hydro-Electric may be more willing to defend their share of supply than are the regional electricity companies in England and Wales.

Some suppliers have, however, complained to me that they have had difficulty in competing in Scotland as a result of their inability to secure access to generation at prices that make them competitive with the local suppliers. I mentioned earlier that present arrangements for second-tier supply in Scotland make use of the Pool price as a reference price for generators and suppliers. The Pool price was intended to act as a ceiling and, in practice, the price at which ScottishPower and Hydro-Electric have been prepared to offer generation to other suppliers has not been below that level. If offers to customers based upon Pool price are not competitive in Scotland, whereas they are in England and Wales, then it suggests that prices to customers in Scotland are already below those in England and Wales. Some suppliers have alleged that their difficulties in securing customers in Scotland must then be the result of discrimination by the Scottish companies between customers and suppliers. ScottishPower and Hydro-Electric have, in turn, claimed that they make offers to customers by reference to their expectation of Pool price. The use of an objective measure, such as Pool price, has some advantage over reliance on a

© Scottish Economic Society 1996

non-discrimination rule as a basis for making available generation to other suppliers. Nonetheless, this is a matter I continue to keep under close review.

Is it inevitable that there should be separate markets and different market prices in Scotland and England and Wales? Inan unconstrained market, output would tend to flow from the lower price market to the higher price market, thereby tending to equalise prices. At present the limited capacity of the interconnector is a constraint on this. However, the Scottish companies recently announced their intention to upgrade the interconnector to 2200 MW from its present 1600 MW. This would represent a near trebling in capacity since privatisation when the interconnector was only 850 MW. It would go a long way to reducing transmission constraints on economic trade between the two countries. The development of two-way trade between England and Wales and Scotland would improve the competitive nature of the industry north and south of the border. It would also increase the convergence of prices in the two countries and, increasingly, make it possible to think in terms of a GB market for electricity. I should add, however, that increasing the effective capacity of the interconnector may depend on construction of NGC's new transmission line over the North York moors. Consent for this new transmission line has been, and still is, the subject of Public Enquiry. A number of environmental objections have been raised to its construction and the Enquiry will allow these to be properly considered against the case made by NGC.

In principle, at least, it would be possible for second-tier suppliers wishing to secure customers in Scotland to do so by importing generation across the interconnector from England and Wales, against the prevailing flow of current. This has not yet been a feature of arrangements in Scotland, but the possibility is being explored.

V PRICE REGULATION

Having examined the pattern of competition in supply, I would now like to turn to those customers who have not yet had the opportunity to benefit from choice in supply but instead have been protected by price regulation since privatisation. The two Scottish electricity companies are subject to an extensive system of price controls. Separate controls apply to their transmission, distribution and supply businesses. In contrast to the situation in England and Wales, the price controls that were set by the Government at privatisation required the Scottish companies to make a reduction in prices year on year in real terms, albeit a modest reduction. Whether these price controls were in effect tougher than those on the RECs and NGC depends on many factors, including the relative starting points, future capital expenditure requirements and scope for increased efficiency, as well as forecasts of demand. Nonetheless, it would appear that ScottishPower's and Hydro-Electric's initial price controls proved to be somewhat tougher than those faced by the companies in England and Wales, and that there was consequently less scope for tightening them in the price control reviews I have subsequently conducted.

With the publication of my proposals for revised supply and distribution

© Scottish Economic Society 1996

price controls in September, I have now proposed revisions to all three price controls for both Scottish companies (Figure 6).

In 1993 I conducted a review of ScottishPower's and Hydro-Electric's transmission price control. Following this, I proposed a tightening of the price control for Hydro-Electric increasing the X-factor from 0·5 to 1·5. I maintained the existing X factor of 1 that already applied to ScottishPower, which was the tightest of the X factors of any price control at privatisation. In contrast, I proposed a tightening from X = 0 to X = 3 for NGC's transmission control.

The supply and distribution price controls which I reviewed last year have a more significant impact on the prices paid by customers. In distribution, I proposed a tightening for both companies, raising ScottishPower's X from 0·5 to 2, and Hydro-Electric's from 0·3 to 1. This contrasts with proposals for the RECs that embodied an X factor of 2 in addition to one-off cuts in allowed distribution charges of between 11% and 17% for those companies. The proposals for the RECs represents a very substantial reversal from the average increase in prices of 1·3% above the rate of inflation allowed to the RECs under their initial price controls. Moreover, as you will be aware, I recently announced that I am considering a further tightening of the REC price controls.[2] I and others have commented on the REC share price increases since publication of my proposals in August. In contrast, the share prices of the Scottish companies

Transmission	• NGC	RPI - 0 →	RPI - 3
	• SP	RPI - 1 →	RPI - 1
	• HE	RPI - 0.5 →	RPI - 1.5
Supply	• RECs	RPI - 0 →	RPI - 2
	• SP	RPI - 0.5 →	RPI - 2
	• HE	RPI - 0.3 →	RPI - 2
Distribution	• RECs*	RPI + 1.3 →	RPI - 6.5
	• SP	RPI - 0.5 →	RPI - 2
	• HE**	RPI - 0.3 →	RPI - 1

Notes: * Based on original proposals.

**Based on original proposals - MMC recommended X be increased to 2 after a further year at 0.3.

Figure 6. Price controls tightened.

[2] Revised proposals for the REC distribution price control were published in July 1995. Taken together with the earlier proposals they imply a price control for the five years 1995–2000 of RPI – 9·5 for the average REC. 'The Distribution Price Control: Revised Proposals July 1995' (OFFER, 1995).

© Scottish Economic Society 1996

have not risen since I published price control proposals for them—indeed they have fallen in absolute terms and relative to the market since then.

As already discussed, the form of the supply price control in Scotland is appreciably different from that in England and Wales, having a more direct impact on the prices that can be charged for generation. It was envisaged at privatisation that from 1994 the prices charged for generation in the price controlled market in Scotland would move gradually to a yardstick based on prices in England and Wales. My review endorsed that transition. As a result I estimated that generation prices to franchise customers would fall by up to 15% in real terms over the next three years, bringing them more closely into line with the level of prices in England and Wales. The impact this may have on the profitability of ScottishPower and Hydro-Electric will be offset to some extent by a parallel reduction in prices charged by Scottish Nuclear where prices also converge on a yardstick related to prices in England and Wales.

I also tightened the price control on the revenue to cover the costs of the supply business itself. The previous X factors were 0·3 for Hydro-Electric and 0·5 for ScottishPower. I proposed to increase both of these to 2.

I estimated at the time that the total effect of my proposals on distribution and supply (including generation) would be to reduce prices to Scottish customers in the franchise market by around 10% in real terms over the next three years. These proposals were accepted by ScottishPower but not by Hydro-Electric. This led to my first reference to the Monopolies and Mergers Commission. The Commission is due to report its findings in the middle of May.[3] Its report will make recommendations on appropriate distribution and supply price controls for Hydro-Electric and will also discuss the 'Hydro Benefit' provision. This provision transfers some of the benefits of the relatively cheap hydro generation owned by Hydro-Electric to its transmission and distribution businesses. At privatisation the Government was concerned that without such a transfer, prices to Hydro-Electric's customers might otherwise be higher than those elsewhere in Scotland, and England and Wales, by virtue of the territory in which Hydro-Electric operates.[4]

VI Prices to Franchise Customers

Let us now look at the overall experience of prices to franchise customers and the implications of the new controls. There has been a pattern of falling prices to customers in Scotland. Since privatisation those customers in the franchise market up to 1994, that is with maximum demand below 1 MW, have seen their prices fall by an average of 7·5% in real terms. This reduction is slightly greater than the equivalent reduction in England and Wales.

Domestic customers are the largest customer grouping within the franchise market. Their price reductions have been slightly lower in Scotland at about 4%

[3] 'Scottish Hydro-Electric plc, A report on a reference under section 12 of the Electricity Act 1989' (HMSO, 1995).

[4] In its report, the MMC endorsed the principle of Hydro Benefit and clarified its method of calculation and application. *op. cit.* pp. 28–32

© Scottish Economic Society 1996

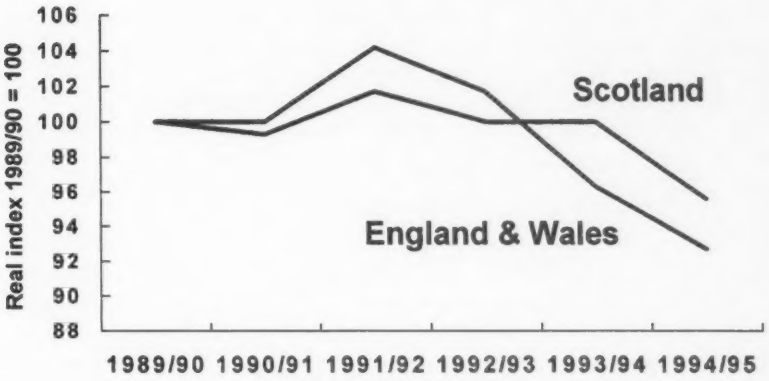


Figure 7. Domestic price reductions.

compared to a fall of domestic prices on average of about 7% in England and Wales (Figure 7). Historically domestic prices in Scotland have been amongst the lowest in Great Britain and ScottishPower and Hydro-Electric remain in the bottom half for domestic prices.

The trend in Scotland of domestic prices has also been different. In England and Wales there was a significant price increase in the second year, before more recent and more than offsetting price reductions. In Scotland there have been steady and then falling prices for domestic customers. In part this reflects the tighter price controls, but to some degree it also reflects the effect of higher than forecast inflation in the first year of the price controls in England and Wales.

VII Full Competition in Supply

Turning to the future, the most significant development over the next few years will be the introduction in 1998 of competition for all customers. In February this year I issued a consultation paper on arrangements for 1998 and have subsequently established two coordination committees, one for Scotland and one for England and Wales, to advise me on all aspects of '1998'.[5]

The different industry structure and pattern of regulation in Scotland, and differences in the evolution of competition to date, make for some potentially interesting discussion as these 1998 arrangements are developed. It will, for example, be important to consider whether the present low penetration of second-tier suppliers is likely to continue after 1998 and, if so, whether that should be a matter for concern and whether further regulatory changes may be needed to encourage the entry of second-tier suppliers. It is also a matter for consideration just how the present arrangements for second-tier supply might need to be modified to facilitate the substantial increase in the size of the

[5] In August 1995, the Director General issued a further statement: 'The Competitive Electricity Market from 1998: The Next Steps' (OFFER, 1995).

© Scottish Economic Society 1996

competitive market after 1998. Discussion of these issues is at a relatively early stage but ScottishPower and Hydro-Electric will be reporting to me by the end of the month with their preliminary thoughts on how the present arrangements might be modified in Scotland. Whatever the particular form those arrangements take, I am committed to ensuring that customers throughout Great Britain enjoy the full benefits that a choice of supplier can bring to them.

VIII Conclusions

In the light of the foregoing review of experience in Scotland, I want finally to return to the question of industry structure. Vertical separation has been a key feature of the privatised industry in England and Wales. The Northern Ireland system, although much smaller than Scotland's, is also characterised by separation of generation from transmission. Other countries that have embarked upon a process of privatisation and introduction of competition have also tended to reconsider the industry structure so as to separate generation and transmission from distribution and supply as in England and Wales. There may, therefore, seem to be something rather anomalous about the industry structure in Scotland.

Five years ago, in my Prospectus statement for the flotation of the two Scottish companies,[6] I noted that the structure and circumstances of the industry in Scotland provide a different background to that in England and Wales for the development of competition and that a number of features of the regulatory regime take account of that background. Whilst I considered that the regulatory regime was capable of ensuring that customers have access to the full benefits of competition, I said that this was something that I would want to keep under review. More recently I have mentioned the possible advantage of greater separation of competitive and non-competitive parts of the industry.

Experience to date suggests that, in terms of price, customers in Scotland have benefited from electricity privatisation at least as much as customers in England and Wales, and in some respects more. However, it remains an open question whether this will continue and, in particular, whether the extent of competition between two adjacent and vertically integrated companies can continue to protect customers as effectively as in a more disaggregated market. As noted, there have been some complaints from other suppliers. Moreover, the Scottish system is as yet untested in certain key respects, such as the ability to cope with the requirements of competition in generation, with a multiplicity of competitors in the supply market, and with importing as opposed to exporting through the interconnector. It is also an interesting question how far the Scottish experience has been dependent on the more radical changes made in England and Wales.

If the charge is that the structure of the industry in Scotland is not conducive to effective competition, at this stage the verdict, appropriately for Scotland, would seem to be 'not proven'. As I forecast five years ago, the regime in

[6] 'The two Scottish electricity companies share offers' (BZW, 1991, pp. 67–71).

© Scottish Economic Society 1996

Scotland has been capable of ensuring that customers have access to the benefits of competition as these have emerged to date. But developments over the coming years will focus increasing attention on arrangements in Scotland, and it will be important to continue to keep these under review. As I have already indicated I am committed to ensuring that Scottish customers continue to enjoy the benefit of private ownership and competition.

Date of receipt of final manuscript: 10 October 1995

THE SCOTTISH DOCTORAL PROGRAMME IN ECONOMICS

The Scottish Universities of Aberdeen, Dundee, Edinburgh, Glasgow, Heriot-Watt, St Andrews, Stirling and Strathclyde combine to offer a joint MSc and PhD in Economics. Training on the MSc extends over a period of 11 months and comprises advanced coursework and completion of a long essay. A further two years research and training are undertaken by those going on to the PhD.

During the first year of the Programme students take compulsory courses in Advanced Macroeconomics, Advanced Microeconomics and Quantitative Methods, together with two optional courses from a range including Labour Economics, International Money and Finance, Planning and Reform in Eastern Europe and Industrial Organisation. A good pass in the coursework allows students to proceed to write the long essay over the period June to August at the end of the first year. Successful completion of the long essay results in the award of an MSc and, for those who achieve the appropriate standard, transfer to the PhD with supervision by at least two members of staff of the universities who are recognised as experts in their chosen field.

The first year of the Programme is taught in Glasgow, those proceeding to the PhD attend one of the eight participating Scottish universities.

Scholarships to support students studying for a PhD are available (under Mode A) from the ESRC. Support is also available from the ESRC under the Advanced Coursework Awards to support students studying for the MSc. The Scottish Economic society and the Carnegie Trust for the Universities of Scotland also provide scholarships to support students on the Programme.

For further details and application form send to:
The Scottish Doctoral Programme & Masters Degree in Economics
Adam Smith Building
University of Glasgow
Glasgow G12 8RT
Scotland
Tel: 0141 339 8855 Ext 4658
FAX: 0141 330 4940

© Scottish Economic Society 1996

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

FINANCIAL INTERMEDIATION AND FINANCIAL INNOVATION IN A CHARACTERISTICS FRAMEWORK

*David Blake**

ABSTRACT

This paper shows that the characteristics model provides a unifying framework for analysing both financial intermediation and the process of financial innovation. The role of the financial intermediary is to buy-in the primary liabilities of the firm, unbundle and repackage the characteristics contained in them and issue a set of ultimate assets to households. Assets exist because the balance between the supply of and demand for asset characteristics indicates an interior equilibrium. Financial innovations arise whenever changes in supply and demand induce movements from a corner solution to the interior. A number of existing explanations of financial innovation can be expressed using this framework.

> The [financial] innovation process has produced many new instruments in a particular historical sequence. An ideal theory of the process should explain how changes in general economic conditions created specific profit opportunities for new instruments to emerge. It should explain all innovations and the order in which they arose and should relate to a wide range of historical experience. There is no accepted theory of innovations that meets all of these criteria ...
>
> It is certainly possible to describe the innovation process in very general terms. Any financial instrument can be viewed as a combination of characteristics ... Financial innovation, then, can be seen as the process of unbundling and repackaging these characteristics to create new instruments.
>
> The Cross Report (1986, p. 169)

I INTRODUCTION

Any financial system has the following four components: claims, services, delivery systems and organisations. Financial intermediation and financial innovation can be associated with changes in any of these. In this paper, we examine financial intermediation and the process of financial innovation in a

* Department of Economics, Birkbeck College, London, W1P 2LL, UK

characteristics framework.[1] Claims and services are defined in terms of both internal characteristics and external characteristics, namely the delivery systems and organisations that provide them.[2] We examine: the supply of internal characteristics by primary borrowers and the demand for them by ultimate lenders; the role of financial intermediaries in unbundling and repackaging these characteristics; and the role of external factors such as the business cycle and changes in technology, constraints and regulations in influencing this process over time.

The approach that we adopt is an integrated one and examines financial innovations from both the demand-side and supply-side. We argue that claims and services exist because the balance between the supply of and demand for their characteristics (both internal and external) indicates an interior equilibrium, and financial innovations arise whenever changes in supply or demand induce movements from a corner solution to the interior.

Conventional explanations have typically concentrated either on the supply-side of the innovation process or on the demand-side. Examples of the former range from those that regard exogenous technical change as the dominant factor inducing innovation, to those that view innovations as being endogenous to the financial system itself, in the one case responding to constraints (or transactions costs) of various kinds (e.g. Silber, 1975, 1983; Ross, 1989), in the other responding to regulations (or taxes) of various kinds (e.g. Kane, 1981, 1983; Miller, 1986). Examples of the latter include Gurley and Shaw (1955, 1956), Silber (1981) and Miller (1991). We demonstrate how all these explanations fit into the characteristics framework and therefore how the characteristics model provides a unifying framework for analysing the process of financial innovation.

In Section II of the paper, we examine financial intermediation in a

[1]The characteristics model of the demand for related goods was originally developed by Gorman (1980) (see also Lancaster, 1966) and first applied to portfolio behaviour by Tobin (1958, 1966, u.d.) and to financial innovations by Greenbaum and Haywood (1971) (see also Greenbaum and Higgins, 1983). There has recently been renewed interest in the financial applications of the characteristics model. Blake (1990) developed a characteristics framework for analysing portfolio behaviour and asset pricing. Heffernan (1990a) provided a characteristics definition of a financial market, while Heffernan (1990b) analysed financial regulations in a characteristics framework. Similarly, high quality data on financial characteristics has become more widely available recently and this has begun to generate empirical applications, e.g. Desai and Low (1987) and Heffernan (1992). The present paper provides the first integrated analysis of financial intermediation and financial innovation in a characteristics framework.

[2]The distinction between claims and services is often imprecise. While pure financial services exist, many financial services are linked to the underlying assets and are therefore part of the definition of the assets themselves. For example, a current account with a cash dispensing facility combines an asset and a service, giving a different asset from that of a current account without such a facility. In the characteristics model however, the distinction between claims and services is not important. What is important is the configuration of characteristics embodied in the claim or service. In the above example, the current account with cash dispenser provides an additional quantity of a particular characteristic (namely liquidity) compared with that of the current account alone. Throughout the remainder of the paper, the analysis will be conducted mainly in terms of assets. However, the framework can easily accommodate pure services. In a similar way, the distinction between delivery systems and organisations can be imprecise.

© Scottish Economic Society 1996

characteristics framework. In Section III, we develop the characteristics model of financial innovation. Section IV concludes the paper.

II THE CHARACTERISTICS MODEL OF FINANCIAL INTERMEDIATION

Financial intermediation can be characterised very simply. Primary claims are issued by primary borrowers (typically firms) and ultimate claims are held by ultimate lenders (typically households). Intermediating between these two groups is a set of financial institutions producing intermediate claims and services. These financial intermediaries will use a variety of delivery systems to distribute and market their claims and services: branch networks, door-to-door sales forces, telephone, post, card transfers, automated teller machines etc.

Niehans (1983) argues that all financial products are functions of three basic financial services that are immutable: the bringing together of borrowers and lenders, the exchange of money between the present and future (including contingent intertemporal transfers such as insurance), and the execution of a payments (money transmission) system. In providing these services, financial intermediaries perform three resource-using functions: the processing of information, the transformation of maturities and the provision of liquidity, and the transformation of risks (see, e.g., Benston and Smith, 1976; Greenbaum and Higgins, 1983). Each of these functions involves substantial scale economies which together justify the existence of financial intermediaries: individuals could not perform them as efficiently or as cost-effectively acting alone.[3]

The collection and processing of information (e.g., on credit risk) can be more efficiently performed by intermediaries than by individual borrowers and investors and this reduces the search costs of both groups (e.g., Leland and Pyle, 1977; Campbell and Kracaw, 1980; Diamond, 1984; Ramakrishnan and Thakor, 1984; Millon and Thakor, 1985; Boyd and Prescott, 1986). The other two functions of financial intermediaries involve asset transformation. One, maturity transformation, resolves the Hicksian (1939) 'constitutional weakness' of unintermediated financial markets through intermediaries holding the long-term illiquid liabilities of deficit sectors and creating short-term liquid liabilities which the surplus sectors wish to hold as assets. As a result of this particular transformation of asset characteristics, financial intermediaries incur substantial maturity mismatches on their balance sheets which leads to various kinds of

[3] Of course, in a perfect market with no frictions such as transactions or information costs or regulations, financial intermediaries would not exist: individuals could costlessly replicate all the activities of financial intermediaries themselves (e.g. Fama, 1980). But neither would a *process* of financial innovation. Under such conditions, apart from those that arise as a result of exogenous technological change, all assets would have been around since the beginning of time. Only with increasing returns to scale or with certain types of market imperfections (e.g. incomplete markets or informational asymmetries), will there be financial intermediaries and an endogenous process of financial innovation. For an alternative approach based explicitly on transaction cost economics, see Williamson (1981, 1988) who extends the seminal work of Coase (1937). For an analysis of financial innovations when markets are incomplete, see Detemple (1990).

© Scottish Economic Society 1996

risk: price risk, inflation risk, interest rate risk, liquidity risk, exchange rate risk, basis risk and default risk. The intermediary aims to reduce these risks by performing its third function namely risk transformation through diversification, thereby creating an additional transformation of asset characteristics. The intermediaries are compensated for providing these functions through intermediation spreads (i.e. by charging various liquidity and risk premia). Financial intermediaries are also involved in disintermediation activities through the provision of services that earn them fee-based income.

We can develop the characteristics model by starting with a financial institution intermediating between a representative primary borrower and a representative ultimate lender. We will assume in this section that the set of external characteristics (related to delivery systems and organisations) is given, so that financial intermediation is concerned with the transformation of the internal characteristics of assets between the end-users of the financial system. Internal characteristics refer to the characteristics of assets that do not depend on the delivery systems and organisations that provide them, e.g. expected return, risk, degree of liquidity (i.e. the time or cost to liquidate the security), term to maturity, divisibility, brokerage (issue/underwriting/transactions) costs, degree of reversibility (i.e. bid-ask spread), effect on leverage, indenture/covenant restrictions, degree of seniority, degree of voting rights, location (domestic, abroad, off-shore), balance sheet visibility and tax liability.[4]

The objective of the primary borrower (which for simplicity we take to be a firm) is to select the capital structure that maximises the value of the firm (or equivalently minimises its cost of capital). To achieve this objective, it will issue primary liabilities with a particular embodiment of characteristics. Formally, the firm performs the following exercise:[5]

$$\underset{q^s}{\text{Max}}\ v = p^{s'} q^s \quad \text{s.t.} \quad z^s = F^s(q^s) \tag{1}$$

where v is the value of the firm, q^s is the set of N^s primary liabilities comprising the firm's capital structure, p^s is the set of market prices for those liabilities,

[4] While this list of characteristics is quite extensive, one of the principal aims of the characteristics model is to simplify matters by explaining behaviour in terms of as small a number of characteristics as possible. It is possible to argue that there are only two relevant characteristics: expected return and risk. For example, from the borrower's viewpoint, credit or quality rating or any change to gearing affects the cost of issue and thus the expected return. So do brokerage costs. Indenture/covenant restrictions usually increase the borrower's risk. An appropriate maturity period lowers liquidity risk. Similarly, from the leader's viewpoint, the degree of liquidity, marketability and seniority affect his risk, and brokerage costs his return. Even in the case of balance sheet visibility and tax liability, these characteristics are only relevant when specific imperfections in the accounting and tax systems exist and, again, they will affect return and risk. Nevertheless, the fuller list of characteristics is presented in order to demonstrate the generality of the characteristics model.

[5] Note that while (1) does not formally involve a wealth constraint, the firm is constrained from increasing q^s unrestrictedly and hence making v, the value of the firm, infinite. This is because the financial system is constrained by the wealth of the household sector which ultimately owns both the firm and the financial intermediary. The financial system's wealth constraint is $w = v + y$ where w is household wealth and y is the value of the financial intermediary. The financial system's wealth constraint is recognised in (5) below rather than in (1).

© Scottish Economic Society 1996

and z^s is the bundle of K internal characteristics supplied by the primary liabilities as quantified by the relationship $F^s(\ )$.

The first-order conditions for a maximum of (1) with N^s liabilities and K internal characteristics are:

$$p_i^s = \sum_j^K \lambda_j^s f_{ji}^s(q^s), \qquad i = 1, N^s \tag{2}$$

where λ_j^s is the shadow price of the jth characteristic supplied and $f_{ji}^s = \partial z_j^s / \partial q_i^s$ is the marginal quantity of the jth characteristic supplied by the ith liability. Now the N^s-equation system (2) implicitly provides the solution to the optimal capital structure conditional on the set of characteristics supplied.

The objective of the ultimate lender (which for simplicity we take to be a household) is to maximise a utility function defined on a set of characteristics demanded subject to a wealth constraint. To achieve this objective the household will demand ultimate assets with a particular embodiment of characteristics. Formally, the household performs the following exercise:

$$\underset{q^d}{\text{Max}}\ u = u(z^d) \quad \text{s.t.} \quad w = p^{d'} q^d \quad \text{and} \quad z^d = F^d(q^d) \tag{3}$$

where $u(\ )$ is the household utility function, w is household wealth, q^d is the set of N^d ultimate assets desired by the household, p^d is the set of market prices for those assets, z^d is the bundle of K internal characteristics demanded, and $F^d(\ )$ is a relationship that quantifies the characteristics embodied in the ultimate assets.

The first-order conditions for a maximum with N^d assets and K characteristics are:

$$p_i^d = \sum_j^K \lambda_j^d f_{ji}^d(q^d), \qquad i = 1, N^d \tag{4}$$

where $\lambda_j^d = (1/\lambda^d)(\partial u / \partial z_j^d)$ is the shadow price of the jth characteristic demanded (λ^d being the marginal utility of wealth) and $f_{ji}^d = \partial z_j^d / \partial q_i^d$ is the marginal quantity of the jth characteristic contained in the ith asset. Now the N^d-equation system (4) (plus the wealth constraint) implicitly provides the solution to the optimal holdings of assets, conditional on characteristics demanded.

Intermediating between the primary borrower and the ultimate lender is the financial intermediary. The rationale for intermediation lies in the fact that, at existing prices, $z^s \neq z^d$, that is, the combination of internal characteristics embodied in primary liabilities are not those most desired by holders of ultimate assets. Whilst there might exist prices that would equate z^d with z^s even without intermediation, we assume that both household utility and firm value are lower at these prices than under a solution with intermediation.

One role of the financial intermediary is to buy-in the primary liabilities of the firm, unbundle and repackage the characteristics contained in them and issue a set of ultimate assets to households. In short, the intermediary absorbs one set of claims and produces another. This is a costly exercise and constitutes the

© Scottish Economic Society 1996

value-added of the intermediary. The objective of the intermediary is to maximise this value-added or spread:

$$\begin{aligned} &\underset{z^s}{\text{Max}}\ y = p^{d'}q^d - p^{s'}q^s \\ &\text{s.t.}\quad q^d = F(z^s) \quad \text{and} \quad z^s = F^s(q^s) \end{aligned} \tag{5}$$

where y is the intermediary's value-added and $F(\ )$ is the intermediary's 'production function' or asset-characteristics technology that links 'inputs' (characteristics supplied by the firm's liabilities, z^s) to 'outputs' (assets demanded by households, q^d).

The first-order conditions for a maximum with K characteristics are:

$$\sum_j^{N^d} p_j^d f_{ji}(z^s) - \sum_j^{N^s} p_j^s f_{ij}^s (z^s)^{-1} = 0, \qquad i = 1, K \tag{6}$$

where $f_{ji} = \partial q_j^d / \partial z_i^s$ is the change in demand for the jth asset in response to a change in the input of the ith characteristic. Now the k-equation system (6) solves for the financial intermediary's optimal holding of characteristics, z^s, as a function of p^d and p^s. Given these, the intermediary uses the asset-characteristics technology $q^d = F(z^s)$ to determine its optimal production of ultimate assets, q^d, with characteristics, z^s, conditional on prices, p^d and p^s:

$$q^d = F(z^s(p^d, p^s)). \tag{7}$$

The effect of financial intermediation is illustrated in Figure 1 in the case of two characteristics, z_1, (the negative of) risk, and z_2, liquidity. The intermediary buys claims with characteristics z^s, and sells claims with characteristics z^d (that is, with lower risk and greater liquidity), thereby enhancing household welfare. A natural measure of the extent or degree of financial intermediation (DFI_o) is provided by the norm of the difference between z^d and z^s:

$$\text{DFI}_o = \| z^d - z^s \|. \tag{8}$$

Financial intermediaries also engage in activities which do not involve intermediation but which generate fee-income rather than a spread. Examples are investment banks in deals that are not fully underwritten and banks engaged in foreign exchange trading and contingent liabilities such as bankers' acceptances, note issuing facilities and revolving underwriting facilities. These activities involve partial or even complete disintermediation. For example, an investment bank might advise a firm in tailoring the terms of a bond issue to the needs of investors. In return, the bank receives a fee, generally related to the size of the deal, but does not engage in any asset transformation activity itself.

With this type of activity, the intermediary simply advises the firm to design its liabilities q^s so that z^s equals z^d. The firm is advised to perform the following exercise:

$$\underset{q^s}{\text{Max}}\ v = p^{s'}q^s \quad \text{s.t.} \quad z^d = z^s = F^s(q^s). \tag{9}$$

© Scottish Economic Society 1996

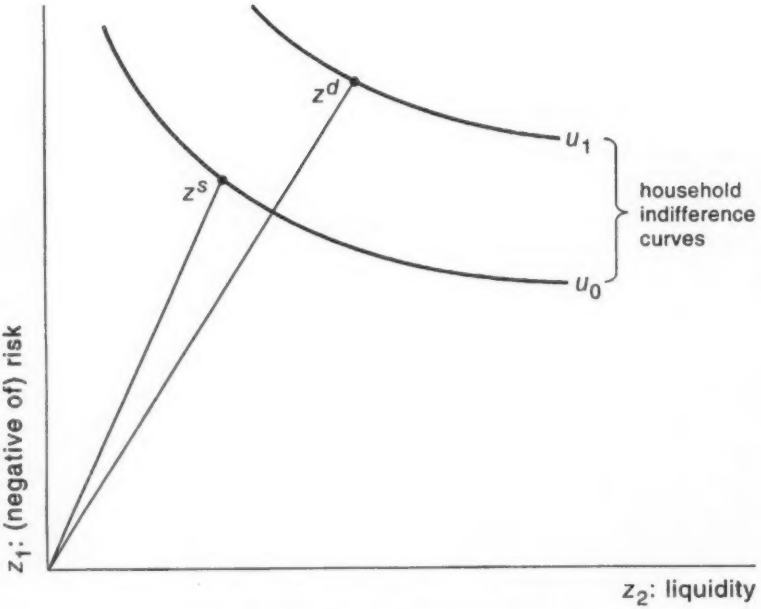


Figure 1. The effect of financial intermediation.

The first-order conditions for a maximum are:

$$p_i^s = \sum_j^K \lambda_j f_{ji}^s(q^s), \qquad i = 1, N^s \tag{10}$$

where λ_j is the shadow price of the jth characteristic (supplied and demanded) and now $f_{ji}^s = \partial z_j^d / \partial q_i^s$. The N^s-equation system (10) solves for the optimal supply of firm's liabilities conditional on the set of characteristics demanded by the household.

The financial intermediary receives a fee (y) which will typically be some proportion γ of the size of the deal:

$$y = \gamma(p^{s'} q^s) \qquad 0 < \gamma < 1. \tag{11}$$

The intermediary and the firm will bargain over the size of γ with the outcome depending on the degree of competition amongst financial intermediaries. The net value of the firm will therefore be:

$$v = (1 - \gamma)(p^{s'} q^s). \tag{12}$$

Both the financial intermediary and the firm have an incentive to maximise the value of the firm, although they will argue over the size of γ. Financial intermediaries will have an incentive to move away from intermediation

© Scottish Economic Society 1996

towards fee-based services when y in (11) exceeds y in (5) (both net of expenses). In the case of financial intermediaries earning fee-income rather than a spread, equation (8) does not provide the appropriate measure of the degree of financial intermediation, since, if the financial intermediary is successful, equation (8) will take the value zero: there is no equivalent to (8) in terms of characteristics in this case.

III THE CHARACTERISTICS MODEL OF FINANCIAL INNOVATION

Although the basic financial services have remained essentially unchanged over time, there exists a process of financial innovation which leads to new types of assets, delivery systems or organisations. Equation (8) measures the degree of financial intermediation and also the first stage of financial innovation. But it does not describe the *process* of financial innovation.[6] For the innovation process to be evolutionary, there must be costs to the introduction of new assets, delivery systems or organisations. Over time, these costs must decline, but at any particular time, they can be substantially fixed and the marginal costs can be low.[7] Alternatively, costless opportunities to innovate could arise exogenously and slowly over time. This is on the supply-side. On the demand-side, there will also be factors such as changing habits (e.g. reducing technology phobia or increasing risk-taking), that lead to innovations occurring as a process over time. Similarly, there are innovations in marketing which can lead to existing products being sold to new client groups, or to apparently 'new' products being marketed when they are really 'old' products suitably repackaged.

To begin with in this section, we hold the set of internal characteristics fixed and consider changes to the set of external characteristics. We then consider enhancing the set of internal characteristics.

[6] Schumpeter (1939) distinguished between invention, innovation and the diffusion of innovations. We concentrate our attention in this paper on innovation. We have little to say about invention, which is essentially a non-economic activity. Nor do we have much to say about the diffusion of innovations, since in the case of financial innovations, competitive or rivalry pressures tend to make the diffusion process a very rapid one (see also footnote 7).

[7] This is true whether the market structure for financial intermediation is perfectly competitive, imperfectly competitive or monopolistic. However, it is likely that the market structure will affect the rate of diffusion of financial innovations. Some (most notably Schumpeter, 1939) have argued that monopoly power is the main source of innovation. Others have argued that it is competition that leads to the fastest development of new products (e.g. Dasgupta and Stiglitz, 1980). Where there is imperfect competition, rivalry is regarded as the main source of innovation (e.g. Kamien and Schwartz, 1975; Scherer, 1967, 1980; Silber, 1975, 1981). Rivalry theory predicts (see Scherer, 1980) that the pace of innovation increases with the number of evenly matched rivals since the potential for the innovator to increase market share is greater. Dominant firms will innovate only when their market share is challenged by smaller rivals, otherwise they have little incentive to innovate in terms of enhanced market share. The conclusion of rivalry theory is that it is the competitive challenge of the rival rather than monopoly power that stimulates innovations. The same conclusion also applies to the rate of diffusion of financial innovations (e.g. Reinganum, 1981). In terms of financial innovations, it is clear that the competitive/rivalry model is the most appropriate for explaining the rate of diffusion.

© Scottish Economic Society 1996

Supply-side innovations

To account for the process of financial innovation from the supply-side, we modify the production function facing the financial intermediary as follows:

$$q_t^d = F(z^s; \theta_t) \tag{13}$$

where θ_t is a set of M external characteristics that defines (or 'constrains') the intermediary and the delivery and information-processing systems that it uses at time t. So (13) defines assets in terms of the internal characteristics that they possess, the intermediary that produces them, the delivery system through which they are marketed and the date at which they are produced. In order to capture these different effects, we will partition θ_t into $\theta_t = \{T_t, C_t, R_t\}$, which respectively, represent the technology, constraints and regulations facing the intermediate at time t. This partition is consistent with existing supply-side explanations of financial innovations, especially those of Silber (*op. cit.*) and Kane (*op. cit.*).

With this modification the optimal production of ultimate assets q_t^d is determined by (cf. (7)):

$$q_t^d = F(z^s(p_t^d, p_t^s); \theta_t). \tag{14}$$

Financial innovations follow when $N_t^d > N_{t-1}^d$ so that $q_t^d = \{q_{t-1}^d, q_{kt}^d\}$ where q_{kt}^d is the new asset (or set of new assets) created during t.[8]

We can define the shadow prices of the external characteristics in terms of their effects on the financial intermediary's value added:

$$\pi_{it} = \frac{\partial y_t}{\partial \theta_{it}} = \sum_j^{N_t^d} p_{jt}^d \psi_{jit}, \qquad i = 1, M \tag{15}$$

where $\psi_{jit} = \partial q_{jt}^d / \partial \theta_{it}$ is the change in demand for the jth asset if there is a marginal increase in the ith external characteristic at time t. For there to be an incentive to innovate at t, it is necessary that $\psi_{kit} > 0$ for *all* potential new assets (q_{kt}^d) and for *some* external characteristic (θ_{it}). Equivalently, the following set of inequalities must hold for *some* θ_{it}:

$$\left| \frac{dq_{jt}^d}{dq_{kt}^d} \right|_{q_{kt}^d = 0}^{dy_t = 0} > \left| \frac{dq_{jt}^d}{dq_{kt}^d} \right|_{q_{kt}^d = 0}^{d\theta_{it} = 0}, \qquad \text{all } k, \text{ all } j \neq k. \tag{16}$$

If the reverse inequality in (16) holds for *all* θ_{it}, then there will be a corner solution and the optimal output of q_{kt}^d will be zero (in addition $\psi_{kit} = 0$, all i). In other words, only if the 'slope' of the objective function (i.e. intermediary's value-added) (evaluated in (q_{jt}^d, q_{kt}^d) space at $q_{kt}^d = 0$) exceeds the 'slope' of at least one 'constraint' will there be an incentive to innovate and hence a movement from a corner solution to an interior equilibrium.

[8] We are associating financial innovations with increases in the number of ultimate assets demanded. It is, of course, possible for financial innovations to be associated with increases in primary liabilities supplied (i.e. $N_t^s > N_{t-1}^s$).

© Scottish Economic Society 1996

While $\psi_{kit} > 0$ and (16) are necessary conditions for an innovation to occur they are not sufficient. This can be seen by looking more closely at the components of θ_t.

Technology-induced innovations: The first component of θ_t is T_t, the set of characteristics defining the technology (e.g. delivery or information-processing systems) available to financial intermediaries. Suppose that before t, the non-availability of a particular technology constituted a binding constraint on the introduction of a new asset, $\psi_{kT,t-1} > 0$ and $q^d_{k,t-1} = 0$. There is an incentive to innovate but no means of doing so. The innovation will not occur until the technology is available at t. If the innovation is introduced in period t, it occurs either exogenously if the new technology is introduced spontaneously, or endogenously if the new technology is developed in response to the lost profit opportunities (i.e. a high shadow price $\psi_{kT,t-1}$) observed in the previous period. Many of the technological improvements in the money transmission system can be explained in this way.

Constraint-induced innovations: The second component of θ_t is C_t, the set of characteristics defining the constraints (either internally (i.e. prudentially) or externally imposed) facing the intermediary. In this case, the incentive to innovate is given by $\psi_{kCt} > \psi^*_{kC} > \psi_{kC,t-1}$. This implies that the cost of abiding by a constraint has been increasing over time and has passed some threshold (ψ^*_{kC}) above which the cost of the constraint in terms of lost profit opportunities has become excessive. Constraint-induced innovations are distinct from technology-induced innovations: in the former case, the constraints can be avoided (either by changing the mix of claims or the delivery system for marketing the claims or the organisational form of the financial institution), but in the latter case the absence of appropriate technology is unavoidable.

Regulation-induced innovations: The final component of θ_t is R_t, the set of characteristics specifying the regulations or the regulatory reaction functions facing the intermediary. Here the problem is more complex than with technology or with constraints because of the reactive nature of regulations, involving a noncooperative dynamic game played between the regulators and the regulated intermediary. Each player's optimal strategy depends on the strategy that the other player is expected to pursue (not only in the current period but in all future periods that the game is played) but does not involve any co-operation with the other player. In the Nash version of the game, each player assumes that the other will attempt to counteract its own action. In the Stackelberg version, the leader moves first and the follower reacts passively.

In deciding its innovation strategy, the financial intermediary must take into account the regulatory authority's objective function. We could motivate this objective function as follows. The regulatory authority wishes to encourage financial innovations and competition between financial intermediaries but has a responsibility for prudential regulation, so is concerned not to allow excess credit, with its attendant inflationary and bankruptcy consequences, to build up in the economy. Historically, the evidence suggests that financial intermediaries have acted as Stackelberg leaders in developing new financial products and the regulatory authorities have behaved like Stackelberg followers, reacting with a

© Scottish Economic Society 1996

lag to the developments in the financial world. But recently, the authorities have been trying to predict what the intermediaries will do in future and take this into account when they revise their regulations. Examples of this have been the 1980 Deposit Institution Deregulation and Monetary Control Act in the US, and the 1980/1987 Banking Acts and 1986 Building Societies Act in the UK.

Using this framework, we can interpret (14), not so much as an optimal decision rule, but as a policy reaction function. A similar reaction function exists for the regulatory authority. For example, the solution in the Nash game (i.e. the subgame-perfect Nash equilibrium (see, e.g. Selten, 1965; Kreps and Wilson, 1982) requires that both reaction functions are optimised across both players and over all time periods simultaneously. Because of this simultaneity, the optimal strategies will be determined implicitly rather than analytically as in (14). The solution will be of the form $\bar{q}^d = \{q_1^d(\bar{R}), \ldots, q_\infty^d(\bar{R})\}$, i.e. the set of the intermediary's optimal strategies (namely, the production of claims from $t = 1, \infty$), and $\bar{R} = \{R_1(\bar{q}^d), \ldots, R_\infty(\bar{q}^d)\}$, i.e. the set of the regulatory authority's optimal strategies (namely, the regulations in force from $t = 1, \infty$).

Again it is possible to explain the process generating new assets in terms of the costs of adhering to the regulations in force. The costs are given by the shadow prices (ψ_{kRt}) in (15). When these shadow prices grow over time and exceed the threshold there is a strong incentive to innovate either by existing institutions or by new institutions established to bypass the existing regulations; and the design of the innovation will anticipate the response of the regulator in current and future periods. Consequently, it is important to differentiate assets according to the institutions that issue them. So, for example, an asset from a regulated deposit-taking institution is different from an asset from an unregulated deposit-taking institution, even though they may have identical internal characteristics. This is because the unregulated institution can alter more easily the characteristics mix of its assets, equivalent to a further innovation, than can the regulated institution in response to changes in regulations by the regulatory authority. However, this does not imply that the regulated institution is in a completely disadvantaged position because it too is able, to a certain extent, to alter the characteristics mix of its assets in a way that substitutes away from the effects of the new regulations. Furthermore, regulations can sometimes work in favour of the regulated intermediary. For example, in the case of deposit insurance and US thrifts, regulation, by enhancing confidence, tends to increase the stock market value of the intermediary in the short-run.

Demand-side innovations

To account for the process of financial innovation on the demand side, we return to equation (4) which provides the implicit demand for assets, given wealth, prices, characteristics etc.

An intermediary introduces a new asset q_k^d with a new combination of internal characteristics. For example, in line with the Cross Report's characterisation of innovations, it could be (price or credit) risk-transferring, liquidity-enhancing,

© Scottish Economic Society 1996

credit-generating or equity-generating. There will be a new asset-characteristics technology $z^d = F(q^d, q_k^d)$ and a new wealth constraint $w = p^d q^d + p_k^d q_k^d$. The resulting demand for the $N^d + 1$ assets is determined by:

$$p_i^d = \sum_j^K \lambda_j^d f_{ji}^d(q^d, q_k^d) \qquad i = 1, N^d + 1. \tag{17}$$

This system of $(N^d + 1)$ equations will in general provide different optimal holdings for q^d compared with (4) because of the new asset-characteristics technology facing investors, $f_{ij}^d(q^d, q_k^d)$, and because of the substitution effects that are induced by the innovation. If we assume that all assets are gross substitutes, then the lower is p_k^d, the issue price of q_k^d, the higher will be the demand for the new asset and the lower will be the demand for existing assets.

In (17), we kept $\lambda_j^d = \lambda_j^d(w)$, the preferences for characteristics, constant. But an important source of demand for financial innovations is likely to be increases in wealth (see, e.g., Gurley and Shaw, 1955). Wealth-induced innovations will result when increases in wealth reduce the shadow prices λ_j^d of demanding characteristics. Similarly, changes in tastes, also reflected in changes in λ_j^d, could lead to the introduction of new assets, delivery systems or organisations. For example, increasing computer literacy has helped the introduction and spread of computer-related financial services such as home banking and share shops.

New internal characteristics

It is also possible for new internal characteristics to be introduced, both on the supply-side (i.e. z^s) and on the demand-side (i.e. z^d). Changes in z^s lead to a new set of inputs into the intermediary's production function while changes in z^d lead to a new asset-characteristics technology. It is clear that financial innovations could be induced by such changes.

In practice, it is rare for completely new characteristics to be discovered since, as Niehans (*op. cit.*) argues, most 'new' characteristics can usually be constructed as a combination of existing ones. However, an extremely important recent example is provided by the introduction of financial options and related derivatives. The most important characteristic of an option is that its price distribution is concentrated on one side of the exercise price. It is not possible (in practice) to replicate this characteristic with any other asset or combination of assets. But its existence now permits hedging strategies that were simply not feasible in the past.

Finally we note that a natural measure of the degree of financial innovation in period t is given by (cf. (8)):

$$DFI_t = \| (z_t^d - z_{t-1}^d) - (z_t^s - z_{t-1}^s) \|. \tag{18}$$

In terms of external characteristics, the degree of financial intermediation will equal $\| \theta_t - \theta_{t-1} \|$.

© Scottish Economic Society 1996

The historical sequence of innovations and innovation waves

Financial innovations have occurred in a particular historical sequence and they also appear to have occurred in definite waves rather than randomly. These phenomena can also be explained in terms of the characteristics framework developed here.

The sequence of innovations can be explained if the rising shadow prices (15) of abiding by existing technology, constraints and regulations pass their thresholds at different times. The sequence of innovations should correspond to the sequence in which these thresholds were triggered. For example, Ben-Horim and Silber (1977), using data on the First National City Bank for the period 1952–1972, found that increases in deposit and capital shadow prices coincided with the introduction of certificates of deposit (in 1961), subordinated debentures (between 1963–1965), eurodollars (in 1969) and bank-related commercial paper and loan repurchases (in 1969).

It is also clear how the sequence of innovations can be bunched together over time. When the shadow price thresholds are triggered, a whole range of new assets with a very similar mix of characteristics will be created at about the same time. While all the innovations are in response to the same threshold being reached (and are therefore essentially identical), they are likely to be produced by competing financial intermediaries. Examples here are: the various types of credit cards; electronic funds transfer systems at the point of sale (EFT/POS); innovations in bond markets such as the different types of coupon-stripped instruments introduced by Salomon Bros (CATS), Lehman Bros (LIONS), Merrill Lynch (TIGRS) and the US Treasury itself on US Treasury bonds (STRIPS) (see also Mason, 1986; Levich, 1988; Matthews, 1994, Table 2-1); and asset securitisation particularly in banking (see Greenbaum, 1987).[9]

Other explanations for innovation waves are the business cycle or the political cycle. If causal factors such as wealth (or income) fluctuate with the business cycle, then we should expect financial innovations to broadly follow the pattern of the business cycle.[10] Changes in governments could also lead to rapid innovation, if the change in government is also associated with a change in regulatory regime.

IV Conclusion

The opportunity for empirical work on financial intermediation and financial innovations using the characteristics model is substantial. Equations (2), (4), (6), (10), (15) and (17) have the form of (nonlinear-in-variables) regression

[9] Even if it could be argued that much of the proliferation in identical new products from competing institutions is the result of non-price competition and the need to preserve market share in the face of first-mover advantage (see, e.g., Anderson and Harris, 1986; Tufano, 1989), it is also possible to explain the ability of such institutions to introduce these products at about the same time in terms of some characteristic shadow price threshold being triggered.

[10] Schumpeter (1939) argues that it is the bunching together of innovations that causes the business cycle.

© Scottish Economic Society 1996

equations. They therefore provide the basis for an empirical test of the validity of the characteristics model, both in time-series and in cross-section form. Assuming that data on variables such as f_{ji}, f_{ji}^s and f_{ji}^d can be collected, testing the importance of the jth characteristic reduces to a test of the significance of the 'regression coefficients' λ_j^d and λ_j^s.

In addition, statistically significant peaks in the time paths of the shadow prices given in (15) should correspond with periods of rapid financial innovation. Prior to an innovation, we would expect to find a statistically insignificant value for the shadow price associated with a particular external characteristic (π_{it}), suggesting that there is no incentive to innovate. Then gradually, the shadow price begins to rise as the cost of not introducing available technology or the cost of abiding by a constraint or a regulation begins to grow. Once this cost has become excessive, an innovation or a series of innovations that use the new technology or that avoid the constraint or regulation will occur. After this stage, the shadow prices decline rapidly to zero again. This is on the supply side. On the demand side, tests will be based on (17). A new asset will only be successful if it contains a significant amount of a desirable characteristic or a significant reduction in an undesirable characteristic relative to existing assets, so that f_{ji}^d is significantly different from zero for *some* j. Testing the significance of the model as a whole, reduces to testing whether the degree of financial intermediation given by (8) or the degree of financial innovation given by (18) is significantly different from zero.

Finally, the characteristics model has a potentially powerful contribution to make in the important new field of security design (see Allen and Gale, 1989, 1994; Harris and Raviv, 1989; Boot and Thakor, 1993; Nachman and Noe, 1994; Duffie and Rahi, 1995).

ACKNOWLEDGEMENTS

The author is extremely grateful for the comments of an anonymous referee and for the advice of Professor L. C. Hunter (the editor).

REFERENCES

ALLEN, F. and GALE, D. (1989). Optimal security design. *Review of Financial Studies*, 1, pp. 229–263.

ALLEN, F. and GALE, D. (1994). *Financial Innovation and Risk Sharing*. MIT Press, Cambridge, MA.

ANDERSON, R. W. and HARRIS, C. J. (1986). A model of innovation with application to new financial products. *Oxford Economic Papers*, (Supplement), 38, pp. 203–218.

BEN-HORIM, M. and SILBER, W. L. (1977). Financial innovation—a linear programming approach. *Journal of Banking and Finance*, 1, pp. 277–296.

BENSTON, G. J. and SMITH, C. W. (1976). A transactions cost approach to the theory of financial intermediation. *Journal of Finance*, 31, pp. 215–231.

BLAKE, D. (1990). Portfolio behaviour and asset pricing in a characteristics framework. *Scottish Journal of Political Economy*, 37, pp. 343–359.

BOOT, A. W. A. and THAKOR, A. V. (1993). Security design. *Journal of Finance*, 48, pp. 1349–1378.

BOYD, J. H. and PRESCOTT, E. C. (1986). Financial intermediary—coalitions. *Journal of Economic Theory*, 38, pp. 211–232.

© Scottish Economic Society 1996

CAMPBELL, T. S. and KRACAW, W. A. (1980). Information production, market signalling, and the theory of financial intermediation. *Journal of Finance*, 35, pp. 863–882.

COASE, R. H. (1937). The nature of the firm. *Economica*, 4, pp. 386–405.

CROSS, S. Y. (Chairman) (1986). *Recent Innovations in International Banking*. Bank for International Settlements, Basle, Switzerland.

DASGUPTA, P. and STIGLITZ, J. (1980). Uncertainty, industrial structure and the speed of R and D. *Bell Journal of Economics*, 11, pp. 1–28.

DESAI, M. and LOW, W. (1987). Measuring the opportunity for product innovation. In M. De Cecco (ed.). *Changing Money*. Blackwell, Oxford.

DETEMPLE, J. B. (1990). Financial innovation, values and volatilities when markets are incomplete'. *Geneva Papers on Risk and Insurance*, 15, pp. 47–54.

DIAMOND, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51, pp. 393–414.

DUFFIE, D. and RAHI, R. (1995). Financial market innovation and security design. *Journal of Economic Theory*, 65, pp. 1–42.

FAMA, E. F. (1980). Banking in the theory of finance. *Journal of Monetary Economics*, 6, pp. 39–58.

GORMAN, W. M. (1980). A possible procedure for analysing quality differentials in the egg market. *Review of Economic Studies*, 47, pp. 843–856 (originally Journal Paper J-3129 of Iowa State College of Agriculture and Mechanical Arts (1956)).

GREENBAUM, S. I. and HAYWOOD, C. F. (1971). Secular changes in the financial services industry. *Journal of Money, Credit and Banking*, 3, pp. 571–589.

GREENBAUM, S. I. and HIGGINS, B. (1983). Financial innovation. In G. J. Benston (ed.), *Financial Services: The Changing Institutions and Government Policy*. Prentice-Hall, Englewood Cliffs, N.J.

GREENBAUM, S. I. (ed.) (1987). Asset securitisation and off-balance sheet risks of depository institutions. *Journal of Banking and Finance*, 11, pp. 355–548.

GURLEY, J. G. and SHAW, E. S. (1955). Financial aspects of economic development. *American Economic Review*, 45, pp. 515–538.

GURLEY, J. G. and SHAW, E. S. (1956). Financial intermediaries and the saving-investment process. *Journal of Finance*, 11, pp. 257–276.

HARRIS, M. and RAVIV, A. (1989). The design of securities. *Journal of Financial Economics*, 24, pp. 255–287.

HEFFERNAN, S. A. (1990a). A characteristics definition of a financial market. *Journal of Banking and Finance*, 14, pp. 583–609.

HEFFERNAN, S. A. (1990b). A characteristics analysis of regulatory changes in the UK financial markets. *Japan and the World Economy*, 2, pp. 107–140.

HEFFERNAN, S. A. (1992). A computation of interest equivalences for non-price characteristics of bank products. *Journal of Money, Credit and Banking*, 24, pp. 162–172.

HICKS, J. R. (1939). *Value and Capital*. Oxford University Press, Oxford.

KAMIEN, M. and SCHWARTZ, N. (1975). Market structure and innovative activity: a survey. *Journal of Economic Literature*, 13, pp. 1–37.

KANE, E. J. (1981). Accelerating inflation, technological innovation and the decreasing effectiveness of banking regulation. *Journal of Finance*, 36, pp. 355–367.

KANE, E. J. (1983). Policy implications of structure changes in financial markets. *American Economic Review*, 73, pp. 96–100.

KREPS, D. M. and WILSON, R. (1982). Sequential equilibrium. *Econometrica*, 50, pp. 863–894.

LANCASTER, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74, pp. 132–157.

LELAND, H. E. and PYLE, D. H. (1977). Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance*, 32, pp. 371–387.

LEVICH, R. M. (1988). Financial innovations in international financial markets. In M. Feldstein (ed.), *The United States and the World Economy*. University of Chicago Press, Chicago.

MASON, R. (1986). *Innovations In The Structures of International Securities*. Credit Suisse First Boston, London.

MATTHEWS, J. (1994). *Struggle and Survival on Wall Street: The Economics of Competition Among Securities Firms*. Oxford University Press, New York.

© Scottish Economic Society 1996

MILLER, M. H. (1986). Financial innovation: the last twenty years and the next. *Journal of Financial and Quantitative Analysis*, 21, pp. 459–471.
MILLER, M. H. (1991). *Financial Innovations and Market Volatility*. Basil Blackwell, Oxford.
MILLON, M. H. and THAKOR, A. V. (1985). Moral hazard and information sharing: a model of financial information gathering agencies. *Journal of Finance*, 40, pp. 1403–1422.
NACHMAN, D. and NOE, T. (1994). Optimal design of securities under asymmetric information. *Review of Financial Studies*, 7, pp. 1–44.
NIEHANS, J. (1983). Financial innovation, multinational banking, and monetary policy. *Journal of Banking and Finance*, 7, pp. 537–551.
RAMAKRISHNAN, R. T. S. and THAKOR, A. V. (1984). Information reliability and a theory of financial intermediation. *Review of Economic Studies*, 51, pp. 415–432.
REINGANUM, J. (1981). Market structure and the diffusion of new technology. *Bell Journal of Economics*, 12, pp. 618–624.
ROSS, S. A. (1989). Institutional markets, financial marketing, and financial innovation. *Journal of Finance*, 64, pp. 541–556.
SCHERER, F. M. (1967). Research and development resource allocation under rivalry. *Quarterly Journal of Economics*, 81, pp. 359–394.
SCHERER, F. M. (1980). *Industrial Market Structure and Economic Performance*. Rand McNally, Chicago.
SCHUMPETER, J. A. (1939). *Business Cycles*. McGraw Hill, New York.
SELTEN, R. (1965). Spieltheoretische behandlung eines oligopolmodells mit nachfragetragheit. *Zeitschrift für Gesamte Stattswissenschaft*, 121, pp. 301–324.
SILBER, W. L. (1975). *Financial Innovations*. D. C. Heath and Co., Lexington.
SILBER, W. L. (1981). Innovation, competition and new contract design in futures markets. *Journal of Futures Markets*, 1, pp. 123–155.
SILBER, W. L. (1983). The process of financial innovation. *American Economic Review*, 73, pp. 89–95.
TOBIN, J. (1958). Liquidity preference as behaviour toward risk. *Review of Economic Studies*, 26, pp. 65–86.
TOBIN, J. (1966). Theory of portfolio selection. In Hahn, F. and Brechling, F. (eds.), *Theory of Interest Rates*. Macmillan, London, pp. 3–51.
TOBIN, J. (u.d.). Properties of assets. Ch. 2 of unpublished manuscript.
TUFANO, P. (1989). Financial innovation and first-mover advantage. *Journal of Financial Economics*, 25, pp. 213–240.
WILLIAMSON, O. E. (1981). The economics of organisation: the transaction cost approach. *American Journal of Sociology*, 87, pp. 548–577.
WILLIAMSON, O. E. (1988). Transaction cost economics. In Schmalensee, R. and Willig, R. (eds.), *Handbook of Industrial Organisation*. North Holland, Amsterdam.

Date of receipt of final manuscript: 20 March 1995

© Scottish Economic Society 1996

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

THE REAL BALANCE EFFECT AND STABILITY ANALYSIS IN CLASSICAL MONETARY THEORY

*Paul Mizen and John R. Presley**

ABSTRACT

The paper investigates the use of the real balance effect mechanism by Dennis Robertson in Money (1922), Banking Policy and the Price Level (1926) and subsequently in his Lectures On Economic Principles (1957–59). The paper draws on the correspondence between Robertson and Don Patinkin during the period 1951–56. It was Patinkin's opinion that Robertson never used the real balance effect as a mechanism to bring the disequilibriated monetary sector back to a new equilibrium; this paper shows that these conclusions are crucially dependent on the interpretation of what constitutes 'stability analysis'.

I INTRODUCTION

During the 1940s and 1950s, Don Patinkin played the major role in the development of neoclassical monetary theory by elaborating upon and highlighting the significance of the real balance effect. The publication of evidence concerning the inconsistency within classical monetary economics, which arose from their dichotomisation of the real and monetary sectors of the economy, helped to establish the need for reconciliation between the two sectors if consistency was to be achieved. Patinkin was not without his challengers, however, over the inconsistency charge: Becker and Baumol (1952) strongly persuaded him to give ground on the recognition of the real balance effect in the works of the classical and neo-classical economists. This he appeared unwilling to do, but he was also being persuaded by other able economists, notably Robertson, who had corresponded with him from May of 1951. The point of this paper is to demonstrate that Dennis Robertson was very instrumental in seeking to persuade Patinkin of the classical and neoclassical understanding of the real balance effect. The focus of this paper is, therefore, the role of Dennis Robertson in debating with Patinkin over the use and interpretation of the real balance effect in the writings of neoclassical monetary theorists. This role,

* University of Nottingham and Loughborough University

whilst characteristic of Robertson's stance on matters of monetary theory, has not been noticed until now because the case was made, to a large extent, in private correspondence. As the paper shows the Robertsonian influence was a significant one.

Section II outlines the classical dichotomy, identified by Patinkin as a mathematically invalid partition of real and monetary sectors of the economy. Section III then deals with the real issue of interest, namely, whether the neoclassicals, and also Robertson, were guilty of Patinkin's charge. It is the changing nature of the charge over the period 1951–56 which is the main point of interest. Patinkin withdraws from the initial attempts to prove the case from the writings of the neoclassicals, particularly Wicksell, to a position of agnosticism based on a lack of clear written evidence in their works. He refers instead to the evidence of the interpretation of his contemporaries and to the use, rather than the existence or interpretation of, the real balance effect in the neoclassical's writings. This final issue concerns the use of the real balance effect for stability analysis which Robertson contends is evident in his own work of the 1920s if not in those of the neoclassicals.

II THE CLASSICAL DICHOTOMY

The invalidity hypothesis

The central proposition and chief innovation in the monetary economics of Don Patinkin was the identification and rectification of the problem of the classical dichotomy. This proposition was outlined in a variety of ways in Patinkin (1948, 1949, 1949–50, 1951a, 1951b, 1956). He demonstrated that the classical system runs into trouble over the determination of the absolute price level because it postulates that the real sector of the economy can be completely specified in terms of relative prices whilst the absolute price level is determined in the monetary sector. The concept of decisions being made in the real sector without reference to the decisions of the monetary sector is implausible, he argued, because money can only be obtained by exchanging goods, and hence the supply of goods is identical to the demand for money. If this is the case, then when the real sector depends on relative prices alone, and not on the absolute price level, the monetary sector must also depend on relative prices alone and the absolute price level cannot therefore be determined.

More formally, if there are n goods and the nth good is money, with prices p_i, $i = 1, 2, \ldots, n$, then the demand and supply functions for money can be specified in terms of relative prices. By making money the numeraire, i.e. setting $p_n = 1$, the demand and supply functions can be written as functions of prices relative to the price of money, in the following way

$$D_i = f_i(p_1, p_2, \ldots, p_{n-1})$$

$$S_i = g_i(p_1, p_2, \ldots, p_{n-1})$$

$$S_i = D_i.$$

© Scottish Economic Society 1996

From this Walrasian system it is possible to demonstrate that the demand for money equals the value of the goods supplied. Patinkin does so by appealing to Walras' law which states that in a general equilibrium system the existence of equilibria in $n-1$ markets implies equilibrium in the nth market, since the excess demands in the n markets sum to zero identically. This provides confirmation of the proposition that the excess demand for money must equal the value of the sum of the excess demands in the other markets.

Money can only be obtained by selling commodities in the other n markets and the supply of money is just sufficient to cover the value of the commodities purchased, therefore:

$$D_n = f_n(p_1, p_2, \ldots, p_{n-1}) = \sum p_i g_i(p_1, p_2, \ldots, p_{n-1}) \quad \text{for } i = 1, 2, \ldots, n-1$$

and

$$S_n = g_n(p_1, p_2, \ldots, p_{n-1}) = \sum p_i f_i(p_1, p_2, \ldots, p_{n-1}) \quad \text{for } i = 1, 2, \ldots, n-1.$$

By subtracting the supply from the demand function the excess demand for money can be written as:

$$D_n - S_n = \sum p_i(S_i - D_i) \quad \text{for } i = 1, 2, \ldots, n-1.$$

Patinkin then demonstrated that the classicals, by virtue of making the assumption that the demand and supply functions are homogeneous of degree zero in the prices $p_1, p_2, \ldots, p_{n-1}$, developed a system which left the absolute price level indeterminate and the real and monetary sectors inconsistent.

Since the original demand and supply equations are homogeneous of degree zero in the prices $p_1, p_2, \ldots, p_{n-1}$, they depend only upon the $n-2$ relative prices $p_1/p_{n-1}, p_2/p_{n-1}, \ldots, p_{n-2}/p_{n-1}$. Likewise the excess demand functions depend only upon the $n-2$ relative prices $p_1/p_{n-1}, p_2/p_{n-1}, \ldots, p_{n-2}/p_{n-1}$, and through Walras' law the excess demand for money function depends only upon these relative prices and not on the absolute price level. The excess demand function for money is homogeneous of degree one.[1]

Patinkin's invalidity hypothesis is then expanded by considering the monetary side of the classical model where the quantity theory equation is used to explain the demand for money in the following way

$$M^S = M = constant$$

$$M^D = kpy$$

where k is an institutional constant, p is the absolute price level and y is the aggregate output of the system. The excess demand function for money is:

$$M^D - M^S = kpy - M.$$

[1] Proof of this point is given in the lemma found in 'The Indeterminacy of Absolute Prices in Classical Economic Theory', p. 14.

© Scottish Economic Society 1996

The inconsistency arises because whilst the excess demand for money function in the real sector was shown to be homogeneous in prices of degree one (such that the absolute price level leaves the equilibrium unchanged) the excess demand function derived from the quantity theory can be shown to be non-homogeneous in prices. Indeed there is a clear effect on the excess demand for money function in the quantity theory relation following a change in the absolute price level, therefore there is an inconsistency between the real and the monetary sectors in the classical model. In the real sector absolute prices are indeterminate and do not affect the excess demand functions whilst in the monetary sector the absolute price level is determinate and excess demand for money is dependent upon it. But through Walras' law the excess demand for money is equal to the value of the excess demands in the real sector and if the latter depend on relative prices alone then the former must depend on relative prices alone.

Put another way it is impossible for the real sector to depend solely on relative prices whilst in the monetary sector the excess demand for money is dependent upon the aggregate price level, as the classical dichotomy suggests. The reason for this being that in a Walrasian system the sectors cannot be held apart since money holding decisions result from the decision to sell goods and therefore the demand for money and the supply of goods decisions are simultaneous. As Patinkin shows, any attempt to partition the sectors results in an invalidity problem.

The real issue however, is not to prove mathematically that there is a problem over the invalidity argument *per se* but rather to prove whether the classicals and neo-classicals were guilty of the invalidity charge in the first place.[2] If, as Robertson argues for the defence, they rectified the alleged inconsistency by appealing to a real balance effect then the charge must be dropped. Patinkin's contribution to monetary theory was precisely this correction to the classical system, making the real side of the model non-homogeneous and ensuring that the real and monetary sides of the model were compatible by including real balances in the excess demand function. The fact that he was at pains to make corrections indicates that he believed that the classical system was invalid without it and did not believe that it had been made beforehand by the neo-classicals or by Robertson. The following section will consider the debate between Patinkin and Robertson to assess whether the classical and neo-classical economists were guilty of omitting the real balance effect in the first place: a charge which Patinkin was prepared to make but which Robertson was not.

[2] Patinkin formalised the real balance effect mathematically in a way which the classicals had not done, but that is not to say that the real balance effect was not recognised by the classical economists. The subject of enquiry here is more to do with a change of emphasis upon the importance of the real balance effect, and it is therefore the recognition of the significance of the real balance effect by the classicals which is under review. The claim that the literature had neglected its influence is not the same as the claim that it was ignorant of it. Nevertheless, the debate between Robertson and Patinkin concerned the ignorance rather than the recognition of the significance of the real balance effect.

© Scottish Economic Society 1996

III THE ROLE OF THE REAL BALANCE EFFECT AND STABILITY ANALYSIS

The Patinkin Pother

The proposition under scrutiny here is whether, as Patinkin argued, the neoclassicals were guilty of an inconsistency in the determination of prices. The conclusions of each party were the result of careful reading, but nevertheless the conclusions were different. Patinkin examined the works of Cassel, Divisia, Walras and Wicksell in particular in the light of this criticism, and drew attention to the passages which he felt supported his objections in their works.[3] Robertson made his response in a letter to Patinkin (20th May, 1951).

Notwithstanding his mathematical illiteracy,[4] Robertson claimed that the Cambridge monetary tradition had always accounted for relative *and* absolute prices through 'the desire to hold command over resources in the form of money being a desire for a real source of utility'[5] that is real money balances (and hence the absolute price level) was a factor in the consumption function.[6]

Patinkin replied that he felt Robertson had 'been too kind to traditional economics'[7] acknowledging that whilst

> it was assumed that the level of real cash balances and hence the absolute price level affects the behaviour of individuals; it is equally clear that in the discussion of the real sector of the economy it was assumed that behaviour was independent of the absolute price level and dependent only on relative prices[8]

Patinkin went on to point out the inconsistency in Wicksell's *Interest and Prices* where absolute prices are ruled out as an important factor on p. 23 but brought in to play a fundamental role on pp. 39–40 with respect to the quantity theory.[9]

[3] 'The Indeterminacy of Absolute Prices in Classical Economic Theory', footnotes 5 and 7; and 'The Invalidity of Classical Monetary Theory', footnote 30.

[4] Letter to Don Patinkin, 20th May, 1951.

[5] *Ibid.*

[6] In a self effacing passage following, Robertson writes that this matter was considered 'in the sort of analysis carried through, clumsily enough, in the algebraic appendix in my "Banking Policy and the Price Level"'. As far as Robertson was concerned it was the pre *General Theory* Keynes who had encouraged him to emphasise the role of the real balance effect in the writing of *Banking Policy and the Price Level* (Robertson, 1926).

[7] Letter to Dennis Robertson, 5th July, 1951.

[8] *Ibid.*

[9] The relevant quote in which Wicksell denies the influence of absolute prices on demand p. 23 '[I]f, on the other hand, the prices of all commodities, or the average price level, is for any reason forced up or depressed there is nothing in the conditions of the *commodity* market that is calculated to bring about a reaction' [italics in original]. However, Patinkin asserts that Wicksell claims that the absolute price level plays an important role in the quantity theory on p. 39 'Now let us suppose that for some reason or other commodity prices rise while the stock of money remains unchanged, or that the stock of money diminished while prices remain temporarily unchanged. The cash balances will gradually appear to be *too small in relation to the new level of prices* ... I therefore seek to enlarge my balance. This can only be done—neglecting for the present the possibility of borrowing etc.—through a *reduction* in my *demand* for goods and services, or through an *increase* in the *supply* of my own commodity ... or through both together' [italics in original].

© Scottish Economic Society 1996

The same inconsistency, he felt, existed in the Cambridge approach but he could not support it with explicit references.[10]

That Robertson disagreed with Patinkin's position with respect to Wicksell is clear; having followed up the references to Wicksell, Robertson wrote

> I cannot see that he [Wicksell] is *inconsistent*. It seems to me he is simply following the well-tried procedure of arguing *first* on an artificially simplified assumption, and *then* introducing the complications. You say that if money is not mere money of account the conclusions of p. 23 need modification; but surely he does give a warning of this straight away, *viz. on p. 24*.[11]

Likewise he claimed that Marshall gave similar warnings of simplifying assumptions.

Patinkin replied ready both to contend and to concede his position (15th August, 1951). Firstly, he conceded to Robertson that Wicksell had modified his conclusions on p. 24 although he disagreed over *which* conclusion had really been modified, noting that once money took on the role of a store of value, the conclusions on p. 23 could not be taken hand-in-hand with the conclusions on pp. 39–40. Where Patinkin was ready to contend for his position, however, was on the issue of whether the modification on p. 24 really made any difference to the question of the classical dichotomy, since the real sector in Wicksell was still dependent on relative prices alone.

Shifting his ground slightly Patinkin conceded that his inferences from the references to Wicksell (and Cassel), in footnote 5 of his paper,[12] concerned the real economy and therefore were not correct. But he took issue with Wicksell and Cassel for not making clear whether the dependence on real prices in the real sector is to be carried over to the monetary sector.[13] It would appear that even as early as 1951 Patinkin began to yield when asked to substantiate his position concerning the dichotomy in classical monetary theory from the writings of the classical and neo-classical writers.

The correspondence between Patinkin and Robertson came to a halt until 1953 when Robertson wrote 'More Notes on the Rate of Interest'[14] which represents a summary of the interchange between Robertson and Patinkin that had taken place in 1951. In 1953 when the correspondence resumed Robertson and Patinkin were joined by Becker and Baumol, who had contributed to the debate (see Becker and Baumol, 1952). Robertson's reading of their article clearly gives the impression that he believed Becker and Baumol were

[10] *Ibid.*

[11] Letter to Don Patinkin 25 July, 1951. The relevant quote from Wicksell p. 24 is 'the exchange of commodities is never in actual practice an instantaneous process, but always extends over some period of time, during which money fulfils the function of a store of value'.

[12] See footnote 4.

[13] Letter to Dennis Robertson 15 August, 1951.

[14] It was in this article that Robertson offered a precis of the correspondence with Patinkin under the sub-title 'The Patinkin Pother'.

© Scottish Economic Society 1996

supporting his line of argument.[15] Patinkin, however, was concerned to ensure that the reading public were not unduly confused by the differences of opinion between himself and Robertson, and especially over the matter of which side Becker and Baumol had taken, since it was his opinion that they had changed their minds.[16] In a letter to Robertson, Patinkin wrote:

> With respect to the classical and neo-classical economists themselves, I believe that once again you are painting with too light a brush. I think the enclosed article[17] demonstrates quite clearly that Becker and Baumol did not grasp the nature of the charge levelled against these economists in the form that it existed at the time they wrote their article. As a result, practically none of their 'excavations' are at all relevant to it. They were, if I may continue your metaphor, digging—without realising it—at a different level than that corresponding to their declared program of activity[18]

and in the following paragraph he wrote

> Frankly, I am not sure that Baumol himself would today accept your statement that his discussion shows that these economists emerged 'without a stain on their characters'.[19,20]

This point is clarified in Patinkin (1954) where he shows that even if Becker and Baumol can prove that the neo-classicals did recognise the real balance effect in an *explicit* way the proof is irrelevant to the line of argument which he is now proposing.[21] Becker and Baumol's argument is valid, he claims, in so much as it demonstrates the irrelevance of the quotations from Wicksell and Cassel previously adduced—although it is notable here that they are considered only 'irrelevant' rather than contradictory to his line of argument.[22] The point which Patinkin wishes to stress is that the neo-classicals did recognise the real balance effect but did so only in the monetary sphere and failed to carry it over to the real sector, therefore, the criticism of the neo-classical economists relates to the *consistency* with which they apply the real balance effect.[23]

At this point there is a further detectable shift of emphasis in the line of argument that Patinkin takes, which he indicates is a consequence of his correspondence with Robertson—he is less concerned to press the charge on the

[15] See 'More Notes on the Rate of Interest' §1 and a letter to Robertson from Patinkin dated 9th October, 1953.

[16] In a letter from Ursula Hicks, then Editor of the *Review of Economic Studies*, to Robertson dated 22nd October, 1953 it is noted that 'Patinkin thinks that he may have made B. and B. change their mind by his attack, [but] we cannot wait until they have signified that in print'. Clearly the matter had not been settled to everyone's satisfaction by this date.

[17] A response to Becker and Baumol's article subsequently published in *Economica*, 1954.

[18] Letter to Robertson 9 October, 1953.

[19] Quoted from the copy of the first section of 'More Notes on the Rate of Interest' sent to him by Robertson.

[20] Letter to Robertson 9 October, 1953.

[21] Patinkin (1954, p. 116).

[22] It is noted on pages 117–18 that the irrelevance of these quotes had been pointed out to Patinkin by Robertson in the summer of 1951.

[23] Patinkin wrote that 'the dichotomy charge had been modified to take it into account' *ibid.* p. 117.

© Scottish Economic Society 1996

classical and neoclassical economists from their own writings, noting that:

> the conclusion expressed here is that the classical and neo-classical writers either did not deal explicitly with the issue at all, or did so in such a vague way that it is impossible either to acquit them completely or convict them completely. The case has been declared a mistrial, with little hope of ever finding conclusive evidence[24]

and

> This has been the main burden of my MS: not to show there was an error, but to show that one could *not* say that 'it is all clearly expounded in the literature'. Thus, my position since the rethinking caused by Robertson's letter is the completely neutral one expressed in Proposition 2: on classical and neoclassical economists, there is neither conclusive positive nor negative evidence. It is only against contemporary economists that evidence exists—and this is uniformly negative.[25]

Rather he is ready to press the argument that writers of 'the literature of the past thirty years',[26] namely Leontief, Haberler, Marget, Rosenstein-Rodan, Hicks and Hickman, have concluded that this dichotomy correctly represents the classical position. So much so that he considered it to be 'the received monetary theory during the past thirty years'.[27] Throughout the exchange of letters it is apparent that Patinkin was less and less concerned to prove his point from the writings of the classical and neo-classical economists and relied instead upon the fact that his fellow economists over the last thirty years had come to accept that the dichotomy was typical of classical and neoclassical monetary theory.

Whilst this argument may have served its purpose by chiding Robertson for confusing the reading public it was surely an erosion of his argument on this actual issue, *viz* the existence of the dichotomy in classical theory; after all if the evidence is vague then the charge can hardly be pressed with such force. But secondly, on what basis did the monetary theorists of the last thirty years have evidence to substantiate their position? and if Patinkin himself could not find sufficient evidence of sufficient weight to support his case from primary sources what weight can be attached to secondary ones? The evidence suggests that Patinkin's argument was evolutionary but the question we are interested in is whether or nor the argument was shifting to firmer ground. On the surface of things the shift of emphasis from primary to secondary sources cannot be interpreted as anything but a weakening of his position.

The correspondence continued until November 1953 but no new issues were raised and the charge concerning the classicals and neo-classicals was no longer mentioned. The final set of letters focused on the related matter of whether the real balance effect had been previously used as 'stability analysis'. The crux of the debate therefore was the interpretation of the term 'stability analysis' or

[24] Letter to Robertson 9 October, 1953.
[25] Letter to Baumol 30 April, 1953.
[26] *Ibid.*
[27] *Ibid.*

© Scottish Economic Society 1996

'dynamic analysis' as Patinkin called it, which describes this process of equilibriation; the invalidity argument was not discussed after this point.

The definition given by Patinkin for the term 'stability (dynamic) analysis' was given as:

> The nature of the market forces which bring the economy to equilibrium from an initial position of disequilibrium.[28]

Patinkin levelled his charges at the neo-classical economists by noting that:

> they frequently remained satisfied with the mechanical comparative statics proposition ... [but] ... they failed to provide a systematic dynamic analysis of the way in which the monetary increase generated real-balance effects in the commodity markets which propelled the economy from the original equilibrium position to its new one[29]

it is precisely this dynamic analysis which was *not* integrated into the Cambridge cash-balance tradition of Marshall, Pigou, Keynes and Robertson.[30]

The nature of stability analysis was developed as Patinkin drew out the theme of his book and the ways in which it improved on neo-classical monetary theory. In response to a doubling of the money supply, his book:

> shows the essence of the quantity theory to lie in the *automatic corrective market forces* which continue to operate through the real-balance effect until this doubled price level is attained.[31]

The crucial issue is whether the real balance effect described by the 'Induced Lacking' mechanism, which Robertson developed from the Cambridge equation by means of his period analysis, constitutes a stabilising mechanism as Patinkin's real balance effect does. In the following the focus of attention will be upon Robertson's publications *Money, Banking Policy and the Price Level* and his *Lectures*.

Robertson did not really turn to monetary theory until he worked on *Money* (1922). His early work on industrial fluctuations focused upon real forces in the trade cycle to the relative neglect of monetary and psychological forces, his textbook *Money* was his first effort at understanding monetary theory and policy in the context of the macroeconomic system. But at this stage he was still learning and did not perhaps reach maturity until the later editions of *Money* and *Banking Policy and the Price Level* (1926). In *Money* (1922), the analysis focused on monetary theory 'as a special case of the general theory of value'.[32] Robertson was keen to bring out two characteristics of the demand for money which makes it different from the demand for other commodities.[33] Firstly, he

[28] *Money, Interest and Prices*, second edition, p. 34.
[29] *Money, Interest and Prices*, second edition, p. 167.
[30] *Ibid.*
[31] *Money, Interest and Prices*, second edition, p. 168, our italics.
[32] Preface to *Money* (1922), first edition, p. vii.
[33] *Ibid.*, pp. 30ff.

© Scottish Economic Society 1996

showed that when an individual loses a unit of money the total utility of the community is not diminished, since the loss represents a transfer of purchasing power to the community at large by increasing the *real* value of their money holdings. But secondly, he showed that if the supply of money as a whole is diminished then in the same way the real value of money balances remaining will be proportionally increased. These aspects of monetary theory led on to an exposition of the quantity theory which were part of the Cambridge oral tradition which traced its ancestry back through Keynes and Pigou to Marshall. It was not until the publication of the third edition, however, in 1928 (after the publication of *Banking Policy and the Price Level*) that Robertson started to consider the real balance effects in detail through his emphasis on the stock of money (through the Cambridge equation) as opposed to the flow theory (through the Quantity Theory). There is a change in the way that the Cambridge equation is treated between these editions which suggests that whilst in 1922 Robertson was prepared to consider K fixed, by 1928 the real balance effect was called into operation which implied that K was now permitted to change in order to restore the original level of real balances, or possibly to attain a new desired level of real balances.

As has already been indicated the amendment to the text of *Money* came after the publication of *Banking Policy and the Price Level*, but it was in the latter in which Robertson first brought the real balance effect into operation.[34] The concepts used are those of Induced Lacking and Automatic Lacking, which refer to the voluntary and involuntary adjustment of money balances respectively, following a monetary disturbance. When new money is injected into the system the forced saving process (Automatic Lacking) and the real balance effect (Induced Lacking) come into operation to bring real balances back to their appropriate level.[35] The purpose Robertson has in mind is to explain the behaviour of the Cambridge K across the cycle. Therefore, it is here in the more detailed exposition of Robertson's monetary theory that the existence of a mechanism to bring the actual money stock back in line with the desired level of money balances is considered. The scenario which Robertson envisages is one where the government disturbs the monetary equilibrium by injecting a quantity of new money.

Robertson and Patinkin also focused on the crucial role of the real balance effect in bringing the monetary sector back to equilibrium from a situation of disequilibrium. The concept of Induced Lacking is introduced (p. 49 of *Banking Policy and the Price Level*) which is the voluntary equivalent to the involuntary Automatic Lacking, that is, the forced saving process. Induced Lacking is a voluntary reduction in consumption out of current output for the purpose of restoring real money balances to an appropriate level, following government intervention which alters the monetary equilibrium. The mechanics of the process are elaborated at great length in the Appendix to Chapter five.

[34] The use of the real balance effect in that place is developed in greater detail in Mizen and Presley (1992).

[35] Which may be the same as, or different from, the real balance prior to the injection or withdrawal.

© Scottish Economic Society 1996

Defining M to be the money stock, H as its real value, K the period of circulation, T to be the output per period and P to be the price level he sets out the initial equilibrium as

$$M/K = P.T/K$$

that is, where the stream of money per period equals the value of the stream of output per period. The price level is therefore by rearrangement

$$P = M/T$$

and the real money stock is

$$H = M/P = T.$$

It is then postulated that the government inflates by X/K units of new money. Robertson then defines S_r to be the total stream of money in any one period, L'_r to be the amount of Automatic Lacking (forced saving) and L''_r to be the amount of Induced Lacking taking place in the corresponding period. Likewise M_r and H_r are the money stock and real money stock in that period.

He goes on to demonstrate the effects of the inflation by looking at the period-by-period equilibria. In the first period he shows that the stream of money, S_1, is:

$$S_1 = M + X/K$$

and the price level, the money stock and the real money stock respectively in that period are

$$P_1 = M + X/T$$

$$M_1 = M + X/K$$

and

$$H_1 = M_1/P_1 = T - (X/K).(1/P).(K - 1).$$

Robertson then shows that the Automatic Lacking taking place in that period will be:

$$L'_1 = X/KP_1 = (XT/K).1/(M + X).$$

Induced Lacking must take place to return real money balances to their original level; in order to do this the real balance equation indicates that $(X/K).(1/P)$ real balances must be withheld from the market for $(K - 1)$ periods. This implies that the Induced Lacking that takes place in the period is

$$L''_1 = (X/KP_1).$$

Robertson takes several periods in turn showing that eventually, after K periods the real money stock is adjusted back to the appropriate level (in this case the original level) and the Induced Lacking ceases. Thus the period analysis demonstrates the behaviour between equilibria and shows the Induced Lacking concept to be the means by which the real balances are restored to equilibrium.

© Scottish Economic Society 1996

The analysis found in the *Lectures* is remarkably consistent with the analysis in *Banking Policy and the Price Level* published thirty years earlier. The use of the Induced Lacking/real balance effect mechanism is prominent in the correction of disequilibria in the monetary sector.

Robertson evidently believed that the omission of the real balance effect in *Money*, which was largely due to its purpose as an undergraduate textbook in the Cambridge Economic Handbook series, was more than made up for in *Banking Policy and the Price Level*. In a letter dated 7 August, 1956 he wrote:

> I don't regard the failure to perform a little stability exercise in a book of the character of 'Money' ... as any evidence against me or my teachers! Personally I should have hoped that there might have been accounted to me for righteousness the elaborate account in 'Banking Policy and the Price Level' (p. 49 and Appl. §1, recapitulated in 'Saving and Hoarding', reprinted in 'Essays in Monetary Theory', p. 78) of how the public in an inflation performs 'Induced Lacking' *to the extent necessary to restore the equilibrium level of its real balances.*[36]

Essentially the point being made by Robertson is that for exposition to students, as in *Money* (1922) and indeed in *Lectures in Economic Principles* (1963, p. 340), he is quite prepared to assume the non-existence of induced lacking in order to simplify the analysis; but for the more sophisticated analysis of *Banking Policy and the Price Level* (1926) and *Essays in Monetary Theory* (1940, p. 78), the stability exercise is included. With respect to the Robertson's first comment, in the above quotation, Patinkin certainly has a point when he says that the stability exercise:

> is indeed a very simple one ... And surely this cannot mean that such an exercise is patently outside the terms of reference of the book 'Money'.[37]

Furthermore, the exercise is only mentioned in *Money* in the third edition (1928) after the publication of *Banking Policy and the Price Level*, but not in the first or second editions (1922, 1924), possibly suggesting that, at the very least, the analysis was not at the front of his mind until Keynes prompted him to think of it during the writing of the latter.

In this sense Patinkin is correct to argue that 'they [the neoclassicals] frequently failed to provide a systematic dynamic analysis of the way in which the monetary increase generated real balance effects in commodity markets'.[38] Robertson frequently failed to provide the systematic analysis, but not always, and the reason for it was the nature of the audience to which he was addressing himself. To make that argument that there was a failure by Marshall, Pigou and Robertson to integrate the dynamic analysis at all, cannot be sustained in every case. The dynamics associated with the real balance effect, which moved the

[36] Letter to Patinkin 7 August, 1956.
[37] Letter to Robertson 17 September, 1956.
[38] *Money, Interest and Prices*, second edition, p. 167.

© Scottish Economic Society 1996

economy from one equilibrium to the next, in the Cambridge cash balance effect can be found in *Banking Policy and the Price Level*. A good proportion of the book was devoted to precisely this.

However, there was a 'general fuzziness from which neoclassical monetary theory suffered as a result of its failure to draw the fundamental distinction between individual experiments on the one hand, and market experiments on the other',[39] in this respect Robertson was unable to dispel the fuzziness perceived by Patinkin.

A further point concerns the reference in the preface to *Money* which notes that the book puts an 'emphasis ... on the theory of money as a special case of the general theory of value'[40] which in *Money, Interest and Prices* is identified by Patinkin as the fundamental reason why the Cambridge cash-balance approach failed to account for stability analysis.[41] By considering value theory and not monetary theory they dealt carefully with the stability of the real sector through the operation of supply and demand exercises but did not consider the exercise for the equilibrium of the absolute price level. Thus, Patinkin claims that the neo-classicals failed to carry over to monetary theory what they were so scrupulous in doing in value theory. However, Robertson's use of period analysis allowed him to use the Cambridge approach to elaborate on the behaviour of money balances and the price level between equilibrium positions. This enabled him to overcome the limitations of comparative statics, which was characteristic of the Cambridge cash balance approach, and to consider the determination and stability of the absolute price level.

The use of the Induced Lacking/real balance effect mechanism in *Banking Policy and the Price Level* to restore equilibrium in the monetary sector following a monetary disturbance has been demonstrated. The final issue is to discuss the distinction or nuance highlighted by Patinkin which he claims distinguished his work from that of the neo-classicals. He wrote that neo-classical theory was criticised:

> not for that element of dynamic analysis which describes the forces propelling the economy toward its new equilibrium position after an initial monetary increase—a problem adequately discussed by many neoclassical economists—but for that aspect which describes the forces stabilising the economy at this new position once it is reached—a problem separated by a nuance from the preceding one, but nevertheless discussed only by Wicksell.[42]

The definition of dynamic analysis drawn here would appear to be somewhat different from the definition presented earlier in the same chapter. The dynamic analysis here refers to forces ensuring stability *about* the new equilibrium whilst the forces he referred to as being omitted in neo-classical theory and present in his book, only one page previously, describe the movement *to* the new

[39] *Money, Interest and Prices*, second edition, p. 171.
[40] See footnote 29.
[41] *Money, Interest and Prices*, second edition, p. 169.
[42] *Money, Interest and Prices*, second edition, p. 168.

© Scottish Economic Society 1996

equilibrium.[43] What exactly Patinkin does mean by dynamic analysis is therefore difficult to determine, and indeed he may have changed his meaning over time.

IV Conclusion

The conclusion, drawn from the correspondence and publications of Robertson and Patinkin, that the charge against the neo-classicals was never firmly established from their own works would, of course, have been uncontested by Robertson and Patinkin but for different reasons. Robertson was unconvinced that the neo-classicals had been guilty of the charge in the first place, whilst Patinkin after a period of consideration following his initial publications on the matter was prepared to conclude that there was insufficient evidence to reach a decision, although this was a far cry from dropping the charges altogether. It is significant that in the course of the debate Patinkin shifted his ground considerably, and that this was a consequence of apparently subtle changes to his argument in the course of the correspondence. Although the changes appear to be changes of emphasis or nuance they are in fact far more important than that, and as we have pointed out, have significant implications. Initially the charge was made from the writings of the neo-classical economists themselves but as time went on the consensus of opinion of monetary theorists of the past thirty years was relied upon more heavily and the evidence from the neo-classicals own works was abandoned. By 1956 Patinkin concluded that none of the neoclassical authors of the Cambridge cash balance equation 'provides as vivid or systematic a picture of the real-balance effect as do Wicksell and Fisher. Nevertheless, they do at various points indicate their recognition of this effect.[44] This is a major change in the argument, albeit that the change takes place by degrees over a period of several years.

With respect to the issue of dynamic analysis in neo-classical writings the conclusion must be that it was not as prominent in their writings as the stability analysis in the real sector, a point which Patinkin stresses with respect to Marshall, Pigou, Nicholson and Chapman.[45] It is the precise logical implications of the real balance effect that Patinkin is concerned with and the omission of these details cannot be refuted. Concerning its treatment in Robertson's works, *Money*, at least in its first two editions, could probably be placed into the same class as the neoclassical writings. However, he appears to have developed the kind of stability analysis—referring to the attainment of a new equilibrium following a monetary disturbance—in some detail under the heading of Induced Lacking in *Banking Policy and the Price Level*, and subsequent editions of *Money*, and the *Lectures* contain explicit reference to the real balance effect as a stabilising mechanism. It is less clear that the second type of dynamic analysis—referring to the maintenance of stability at the new equilibrium was

[43] See quotations above, footnotes 23 and 25.
[44] *Money, Interest and Prices*, Appendix G:1 second edition, p. 604.
[45] *Ibid.*, Appendix G:2 second edition, pp. 604–5.

© Scottish Economic Society 1996

covered in such a way in his works. But again Patinkin would appear to have shifted his ground somewhat in the course of the argument, by redefining stability analysis to mean the analytics of the *maintenance* of an equilibrium rather than the dynamic adjustment process *towards* a new equilibrium from the original.

This paper has highlighted the way in which the correspondence of Patinkin and Robertson led to an examination of the existence, use and scope of the real balance effect in the writings of neoclassical monetary economists. The conclusion we draw is that Patinkin made large steps forward in proving the mathematical inconsistency of a dichotomised economy and brought the solution to this problem—the real balance effect—into sharp focus. However, his charges levelled at the neoclassicals, that they did not recognise the effect, do not appear to have survived Robertson's cross-examination intact; this suggests that the *novelty* of Patinkin's idea of a 'real balance effect' is called seriously into question. The interest lies in observing the counsel for the prosecution realising this in the course of the trial, shifting his ground from the existence, to the interpretation, and lastly the use of, the real balance effect in the neoclassical writings.

ACKNOWLEDGEMENTS

The authors are grateful for research funding from the British Academy and the University of Nottingham Administered Research Fund Grant No. UARF010, and the Wincott Foundation; and to Donald Moggridge, Denis O'Brien, David Laidler, Don Patinkin and two anonymous referees for their comments and suggestions. Errors remain the authors' own responsibility.

REFERENCES

BECKER, G. S. and BAUMOL, W. J. (1952). The classical monetary theory: the outcome of the discussion. *Economica*, 19, pp. 355–376.

MIZEN, P. D. and PRESLEY, J. R. (1994). Buffer stock ideas in the monetary economics of Keynes and Robertson. *History of Political Economy*, 26, 2, Summer, pp. 193–202.

PATINKIN, D. (1948). Relative prices, Say's law and the demand for money. *Econometrica*, 16, pp. 135–54

PATINKIN, D. (1949). The indeterminacy of absolute prices in classical monetary theory. *Econometrica*, 17, pp. 1–27.

PATINKIN, D. (1949–50). A reconsideration of the general equilibrium theory of money. *Review of Economic Studies*, 18, pp. 42–62.

PATINKIN, D. (1951a). The invalidity of classical monetary theory. *Econometrica*, 19, pp. 135–51.

PATINKIN, D. (1951b). Further considerations of the general theory of money. *Review of Economic Studies*, 18, pp. 186–195.

PATINKIN, D. (1954). *Economica*. Dichotomies of the pricing process in economic theory, 21, pp. 113–28.

PATINKIN, D. (1956). *Money, Interest and Prices*, first edition. Harper and Row (second edition 1965).

PRESLEY, J. R. (1978). *Robertsonian Economics*. London: Macmillan.

PRESLEY, J. R. (1986). Keynes and the real balance effect. *The Manchester School*, 54, pp. 22–30.

© Scottish Economic Society 1996

PRESLEY, J. R. (1989). J. M. Keynes and D. H. Robertson: three phases of collaboration. *Research in the History of Economic Thought and Methodology*, 6, pp. 31–46.

ROBERTSON, D. H. (1922). *Money*, first edition. Cambridge Economic Handbooks, Nisbet and Son (second edition 1924, third edition 1928).

ROBERTSON, D. H. (1926). *Banking Policy and the Price Level*. Westminster: P. S. King and Son.

ROBERTSON, D. H. (1956). More notes on the rate of interest. *Review of Economic Studies*, 55 (1953–4) and reprinted in *Essays in Money and Interest*. Fontana Library.

WICKSELL, K. (1936). *Interest and Prices*. Macmillan, London.

Date of receipt of final manuscript: 26 May 1995

© Scottish Economic Society 1996

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

SIR JAMES STEUART'S STATESMAN REVISITED IN LIGHT OF THE CONTINENTAL INFLUENCE[1]

*Deborah Redman**

ABSTRACT

The author assesses the influence Continental thought had on Sir James Steuart and places him in the Continental setting in order to understand his most controversial message, his theory of the state, and how it differed from that of his contemporaries, above all from Adam Smith's.

I INTRODUCTION

The glaring differences between Steuart and Smith's works—Steuart's advocacy of broad state intervention and his Keynesian welfare-state lineage—cannot be evaluated properly without an understanding of the impact of Continental thought, especially French and German ideas, on his theory. Assessing the influence Continental thinkers had on Sir James Steuart and placing him within a Continental setting will aid us in grasping what was to become his most controversial message, his theory of the state, and how it differed from that of his contemporaries, above all from Adam Smith's.

That Continental thinkers were an inspiration for Steuart's *An Inquiry into the Principles of Political Oeconomy: Being an Essay on the Science of Domestic Policy in Free Nations* (1767, revised in 1805) is indisputable. In the preface to his *Principles* Steuart (1966, pp. 3–4) described his work as the cumulation of 'the successive labour of many years spent in travelling' and admitted that the 'modes of thinking, also, peculiar to the several countries where I have lived, have, no doubt, had an influence on what I have written' (1966, p. 4, preface).

[1] An abridged version of this essay originally appeared in German as 'Die Figur des Staatsmanns bei Sir James Steuart im Lichte des kontinentalen Einflusses,' with J. Starbatty, in *Über Steuarts »Principles of Political Economy«*, pp. 81–94. Klassiker der Nationalökonomie series, edited by Wolfram Engels, Herbert Fox, Friedrich A. von Hayek, Horst Claus Recktenwald, and Bertram Schefold, Düsseldorf: Verlag Wirtschaft und Finanzen, 1993. Parts of this essay are reproduced here by permission of the Verlagsgruppe Handelsblatt.

* George Washington University, Washington DC.

Although the reviews of the *Principles* were on balance good, the British reviewers complained of the un-English quality of the work (see Steuart, 1966, p. 4, preface).[2] Echoing this, Schumpeter (1954, p. 176, n. 9) observed that 'there is something un-English (which is not merely Scottish) about his views and mode of presentation.' It is to this un-English, Continental component of Steuart's work and its ramifications that this paper is dedicated. Marx (in Sen, 1957, p. 22), however, even went so far as to contend that Steuart's work was 'contrary to the spirit of the eighteenth century'—a view that this paper will take to task.

For Steuart the 'un-English' quality of his work meant its international flavour and impartisan nature. The *Principles*, he insisted in the preface (1966, p. 4),

> will not, in general, correspond to the meridian of national opinions any where; and of this it is proper the reader should be apprised, that he may not apply to the domestic circumstances of his own country what was intended to refer to those of other nations. ...

Indeed Steuart was convinced his stay on the Continent had made his work more impartial than that of his contemporaries:

> The wandering and independent life I have led may naturally have set me free, in some measure, from strong attachments to popular opinions. This may be called impartiality (1966, p. 10, preface).

His view matches exactly Montesquieu's sentiments in *The Spirit of the Laws* (first published in 1748), where Montesquieu emphasized that his principles reflected not his prejudices, but 'the nature of things' (in Richter, 1977, p. 170).

Montesquieu and Steuart clearly felt themselves citizens of the world, in their view a requisite neutrality condition believed to guarantee the scientific validity of their work.[3] Steuart in fact opens the preface to his *Principles* (1966, p. 3) with this observation:

> It is with great diffidence that I present to the public this attempt towards reducing the principles, and forming into a regular science, the complicated interests of domestic policy.

This disinterestedness, or neutral international perspective, was in Steuart's view acquired consciously during his travels and represented the proper attitude

[2] Steuart, however, thought it ridiculous that his ideas be viewed as purely English. 'Can it be supposed,' wrote an annoyed Steuart in response to the reviewers' criticism of his international perspective, 'that during an absence of near twenty years, I should in my studies have all the while been modelling my speculations upon the standard of English notions' (1966, p. 4). It is therefore most curious that Johnson (1960, p. 234) should contend that in 'Steuart's *Political Oeconomy* is to be found the best simulation of *pre-Smithian British economic theory*' (emphasis added).

[3] Montesquieu had visited Austria, England, Germany, Holland, Hungary, and Italy and regarded 'all the peoples of Europe with the same impartiality as I do peoples of Madagascar' (in Richter, 1977, p. 15).

© Scottish Economic Society 1996

to be assumed by both the governmental authority (which Steuart calls the 'statesman') and the political economist.

Although Steuart (1713–1780) was Adam Smith's (1723–1790) contemporary, his *Inquiry into the Principles of Political Economy* takes a fundamentally different approach than Smith's *Inquiry into the Wealth of Nations*.[4] The role of government envisioned by Steuart and the related uncanny family resemblance with Keynes are the two things most sharply distinguishing Steuart's thought from Smith's. Sen (1957, p. 185) was right to maintain that the 'real comparison between Steuart and Smith is, then, a comparison between their fundamentally different approach to economic problems,' if by approach he means not method but a way of viewing the workings of the economy. Skinner has argued (1981, p. 20; in Blaug, 1991, p. 274) that 'Steuart's great experience of European conditions helps to explain his preoccupation with economic policy.' But more than a preoccupation with economic policy sets Steuart's contributions apart from his contemporaries. We will see that this different approach stems chiefly from the German and French influence, especially that of the German cameralists (Justi in particular) and Montesquieu. But before delving more deeply into these differences, we will first review his experiences on the Continent and assess the common ground he and Smith share.

II THE CONTINENTAL EXPERIENCE

Steuart spent 22 years abroad, between the ages of 22 and 50, five of which were voluntary. After having qualified as a Scottish advocate in 1735, Steuart debarked on a grand tour of Europe to familiarize himself with the customs, manners, and habits of other cultures, a customary undertaking for the gentlemen of his day. During this 5-year tour (1735–1740) he visited France, Germany, Holland, Italy, and Spain and acquired a knowledge of four foreign languages (French, German, Spanish, and Italian). While in France, Spain, and Italy he met with Jacobite leaders, contacts that were later to seal his fate of exile.

In 1740 he returned to Edinburgh, gave up the roving lifestyle, settled in, and married two years later. Five years later the Jacobite leader, Prince Charles, whom Steuart had met in Rome, captured Edinburgh, courted and won Steuart's support. But in 1746 the Jacobite defeat at the battle of Culloden Moor resulted in Steuart's 17-year exile from Scotland.

When his banishment became imminent James Steuart was in France, where he and his family remained for seven years and where Steuart in 1749 initiated his studies in political economy. With the outbreak of the Seven Years' War approaching, Steuart moved his family to Frankfurt am Main, where his first publication appeared, *Apologie du Sentiment de Monsieur le Chevalier Newton sur l'ancienne chronologie des Grècs, Vindication of Newton's Chronology*

[4] For the remainder of this paper Steuart's work will be referred to in abbreviated form as the *Principles* and Smith's as the *Wealth of Nations*.

© Scottish Economic Society 1996

(1757). To be able to provide his son and only child with a good education the Steuarts moved in 1757 to Tübingen, where the family spent the four happiest years of their exile.[5] There his first publication on economics appeared, 'A Dissertation upon the Doctrine and Principles of Money, applied to the German Coin' (1761), and Steuart finished the first three books of his *Principles* (no small task since it comprised five books and 1,284 pages in completed form).[6] Steuart ostensibly spent his days writing the *Principles* and his evenings in the company of various University of Tübingen professors.[7] Although Taylor (1957, p. 298; in Blaug, 1991, p. 96) concludes that their 'chief pastime seems to have been the discussion and close following of the fortunes in the war in Europe,' it is unlikely that Steuart would have neglected discussing his chief diversion, the *Principles*, with them.

Only in 1763, after the Treaty of Paris had been concluded, did Steuart return to Scotland, although he had to wait until 1772 to receive an official pardon. At his Coltness estate he immediately commenced writing the last two books of his economic treatise. The *Principles* was published in two imposing tomes in 1767 by the same house that would publish Smith's *Wealth of Nations* nine years later. A point well worth underscoring: Steuart's travels had made him much more knowledgeable about economic conditions in Continental Europe (especially in France and Germany) than in Britain.[8] And the conditions were not comparable: Britain's economy was growing, the economies of the other European states were stagnating—a situation that most possibly explains Steuart's interest in development, growth, and the stages of capitalist evolution.

III Common Ground with Smith

In attempting an assessment of the intensity and expression of the Continental influence, it will be instructive first to establish the similarities between Smith and Steuart's work. Anderson and Tollinson (1984, p. 456) have argued that Steuart 'was virtually the complete antithesis of Smith'—an untenable assertion. Both men authored comprehensive treatises on political economy in

[5] Steuart chose Tübingen for the good reputation of the professors and its distance from the war (Taylor, 1957, p. 298; in Blaug, 1991, p. 96).

[6] The *Principles* occupies four of the six volumes of Steuart's collected works, *The Works, Political, Metaphysical, and Chronological of Sir James Steuart*, whose compilation and publication was seen through in 1805 by Steuart's son. Skinner's abridged two-volume edition of the *Principles*, which omits approximately one-fourth of the original text, still comprises 735 pages.

[7] Cf. the following excerpt from the Coltness Papers (in Skinner's introduction to Steuart, 1966, p. xl): 'His time was pleasingly divided between the prosecution of his studies and the company of the Professors. When the university hours were over the Professors used to repair to the house of Sir James Steuart and the evening was spent in a mutual communication of ideas.'

[8] Skinner (1966, p. lxxxii) reaches a similar conclusion: 'in the last analysis it must be recognised that Steuart knew little of England; that his audience, his experience and his outlook were all European. He could, and did, accuse the reviewers who criticised his lack of emphasis on England, of insularity.'

© Scottish Economic Society 1996

which they, more importantly, used similar methods. Like Cantillon, Montesquieu, and Smith, Steuart employed the Scottish method of conjectural history to analyse the laws and customs of different peoples.[9] This is the practice of tracing all social institutions to their origins and then constructing stadial theories of explanation that compare 'rude' to 'polished' states. Hence in the first book of the *Principles* Steuart tells us in characteristic Scottish style that he 'shall then set out with taking up society in the cradle' (1966, p. 28, I, ii).[10]

A theory of economic evolution underlies both Smith and Steuart's systems of thought. Steuart, like Smith, assumed nations go through a series of stages of economic development, in Steuart's case a savage, agricultural, and manufacturing stage. But here Steuart parts ways with Smith. The exchange economy was in turn characterized by three stages: infant trade, foreign trade, and inland trade. This theory of stages of trade is really at the heart of Steuart's analysis since he used it to show that unequal levels of economic development limit free trade, justifying protective measures. As Chamley (1963, p. 77) remarks, Steuart opened our eyes to the fact that states unequally developed are controlled by power relations. According to Steuart's plan, economic policies were to be geared to each stage of economic development, as well as to the institutions, traditions, and circumstances of each country. The government was to encourage both agriculture and industry, but with a different emphasis for each of the three stages of capitalist development. Tools of control advocated by Steuart include legislation, manipulation of the money supply, taxation, and public expenditure.

In the infant-trade stage Steuart urged the government to encourage industry by introducing machines and manufacturing and insisted on protective measures without which infant industry would not survive. According to Steuart, protection allowed the economy to reach the phase of foreign trade, at which point restrictive policies are no longer necessary. In the foreign trade phase the object of policy was to banish luxury, encourage savings, and hold prices down. Steuart was certain that the boom phase would inevitably come to an end. As foreign trade declines and the nation loses its competitive advantage, the stage of inland trade sets in. Steuart supposed that 'relatively developed nations would oscillate between the stages of foreign and inland trade' (Skinner's introduction to Steuart, 1966, p. lxxviii). He believed economic decay was inevitable but not irreversible: 'no trading state has ever been of long duration, after arriving at a certain height of prosperity' (Steuart, 1966, p. 195, II, x).

The works of the Scottish philosopher-historians were distinguished additionally by a historical and institutional dimension, for they endeavored to

[9] Writes Meek (1958, p. 112; in Blaug, p. 300): 'The most striking fact which emerges from a reading of Steuart is surely his intellectual kinship in this respect with the members of the Scottish Historical School of the eighteenth century.' Skinner (1992, p. 237), however, argues that 'Smith also differs from Steuart and Hume in respect of his use of the historical method. In Smith's hands, the history of civil society is essential to our understanding of the exchange economy and of the social and political environment which it may produce, but it is now the *preface* to political economy rather than integral to the treatment. In any event, Smith did not use the historical method in dealing with economic questions. ...'

[10] The parenthetical bibliographic references provide the respective book and chapter number to facilitate locating the citations in other editions of the *Principles*.

© Scottish Economic Society 1996

support their ideas with a wealth of data, statistics, and details. (Steuart was later to be praised for the accuracy of the details collected on the Continent.)[11] Montesquieu, a personal acquaintance of Steuart's and probably his most important French source of inspiration, had employed conjectural history to examine why different laws and customs exist in various nations. For Montesquieu the answer lay in the 'spirit of the society,' meaning not the German metaphysical concept of a *Volksgeist* but concrete reasons such as varying religion, traditions, economic systems, trade systems, styles of thought, and other institutions.[12] In this respect Steuart borrowed generously from Montesquieu: 'the spirit of the people' and its influence on government is an underlying theme throughout the *Principles*.

There is a further common methodological element: both Smith and Steuart emphasized the necessity of developing a system of economic principles while cautioning against its drawbacks. In the preface to the *Principles* (1966, p. 7) Steuart underscored the nature of his work as 'a deduction of principles, not a collection of institutions,'[13] but warned of the limitations of abstract deductive systems. The object of excess was the French *système*, whose limitations Steuart (1966, p. 8, preface) discussed in the following way:

> These [systems] are no more than a chain of contingent consequents, drawn from a few fundamental maxims, adopted, perhaps, rashly. Such systems are mere conceits; they mislead the understanding, and efface the path to truth. An induction is formed, from whence a conclusion, called a principle, is drawn and defined; but this is no sooner done, than the author extends its influence far beyond the limits of the ideas present to his understanding, when he made his deduction.[14]

Throughout the treatise Steuart repeatedly warned against oversimplification caused by our viewing generalizations 'in a light too confined' and by 'our not attending to the influence of concomitant circumstances, which render general rules of little use' (1966, p. 7, preface).[15] Elsewhere he argued: 'in every part

[11] Skinner (1981, p. 41; in Blaug, 1991, p. 295) relates the anecdote that Dugald Stewart advised his students to begin their studies with the *Wealth of Nations* and then commence to the *Principles*, a work worthy, according to Stewart, for its 'great mass of accurate details' acquired through Steuart's 'personal observation during this long residence on the Continent.'

[12] Steuart's emphasis on institutions, history, and the spirit of the people was later to win him the praise of the German historical school and especially of Wilhelm Roscher, who went on to adopt Steuart's three-stage theory of economic development.

[13] Consider also: 'my intention is to attach myself principally to a clear deduction of principles, and a short application of them to familiar examples, in order to avoid abstraction as much as possible' (Steuart, 1966, p. 18, I, introduction).

[14] Elsewhere Steuart reminds the reader: 'I pretend to form no system ...' (Steuart, 1966, p. 19, I, introduction). Smith (1976b, pp. 233–34) condemned systems along very similar lines and with a wording very close to Steuart's in the *Theory of Moral Sentiments*. Montesquieu also warned against intellectual systems, pointing to the diversity of human institutions, while nonetheless insisting he had set forth first principles derived from history.

[15] At a later point in the *Principles* he writes: 'I am very much for limiting general propositions. I have hardly ever escaped being led into error by every one I have laid down. Nothing is so systematical, nothing so pretty in a treatise as general maxims; they facilitate the distribution of our ideas, and I have never been able to dash them out but with a certain regret' (Steuart, 1966, pp. 67–68, I, xi).

© Scottish Economic Society 1996

of the science of political oeconomy, there is hardly such a thing as a general rule to be laid down' (1966, p. 339, II, xxviii). Steuart (1966, p. 358, II, xxix) reached the conclusion that general propositions 'are only true or false, according as they are understood to be accompanied with certain restrictions, applications and limitations' (see also 1966, p. 4, preface).

Final areas of agreement: both scholars made self-interest and the progress of mankind central to their works. And both use the expression political economy to mean the policy of the state, following French usage, *économie politique*, as initiated by Antoine de Montchrétien (1575–1662). Nonetheless, Steuart's political economy does not neglect the older emphasis on household management of the polity. Finally, the purpose of both of their economic treatises is to show how prosperity is possible in a free society.

IV Deep Differences in Object and Focus

The Lycurgus controversy

There remain the deep differences. The major divergence is often summed up this way: Steuart supposed a managed economy; Smith assumed the economy could manage itself. We might add that it has been fashionable to dismiss the importance of Steuart's 'statesman' and his economics of control, presumably because it has been a source of embarrassment to liberal neoclassical economists. But Steuart's reasons for advocating a strong centralized state are, however, well worth analyzing, not only for their historical significance but for the message for us today.

The source of trouble for many a reader is Steuart's praise for Lycurgus's Sparta in book II, chapter 24 of his *Principles* (1966, pp. 218–27, II, xiv), which opens with this remark: 'The republic of Lycurgus represents the most perfect plan of political economy, in my humble opinion, anywhere to be met with, either in ancient or modern times' (1966, p. 218, II, xiv). This viewpoint led Anderson and Tollison (1984, p. 459) to argue that Steuart's 'purpose in writing the *Principles* was to offer a detailed vision of a rationally planned dictatorship.' They insist, moreover, that Steuart scholars, in particular A. S. Skinner, have 'trivialized' the message of this chapter. But we will see that Anderson and Tollison's interpretation of Steuart's statesman does not stand careful scrutiny. Indeed Steuart's infatuation with Sparta appears at first glance to be a contradiction since it conflicts squarely with Steuart's thesis that the statesman's duty is to channel the good forces in society and counteract the bad. Yet a careful reading of Steuart and his contemporaries' writings reveals several good reasons for Steuart's support of Lycurgus's Sparta.

First, praise for Lycurgus's disciplinary program is common among Steuart's contemporaries, especially Montesquieu, whom Steuart relied on heavily.[16] Much of Book IV of Montesquieu's *Spirit of the Laws* is dedicated to the role

[16] Skinner (1963) also makes the point that Steuart's particular focus can be traced to Montesquieu's *Spirit of the Laws* and his Scottish philosophical bent.

© Scottish Economic Society 1996

of moral education in republics. There Montesquieu followed Aristotle's teaching that a country must instill in all its citizens a love for the state and respect for its laws and that public welfare must come before private welfare. Montesquieu even praises the Spartans for their development of martial skills since they were necessary for the existence of the state (cf. Richter, 1977, p. 79). Historically viewed, Lycurgus's strict military organization and rigid educational system succeeded in transforming Sparta into the prototype polity in which citizens are subordinated to the state. Nonetheless, Montesquieu (in Richter, 1977, p. 202) specifically noted that this disciplinary education is 'peculiar to democracies' and not to be associated with dictatorships. His praise for Spartan discipline notwithstanding, Montesquieu is classified by political scientists as a constitutionalist today; Montesquieu considered himself to be a proponent of moderate limited government, and with Steuart, supported an absolute monarchy. After placing Steuart's work in its historical setting, it is difficult not to gain a sense that in this controversial chapter Steuart is merely echoing Montesquieu.

Moreover, a reading of Steuart that finds the endorsement of a dictorial regime misinterprets the prevailing intellectual climate in the eighteenth century. The concepts of democracy, liberty, and freedom were used in a different sense then than they are in the twentieth-century United States. Steuart's support for the statesman was based on a particular strain of ancient Greek thought with roots in Platonic concepts of freedom and liberty. As Ryan (1987, p. 175) so aptly explains,

> Plato was positively hostile to freedom, which he identified with an unbridled opportunity to do whatever we liked. ... It was not freedom but discipline based on philosophical illumination which would make individuals good and society stable.[17]

Discipline's great virtue was its tendency to contribute to conformity of opinion and thus to discourage dissension. In Steuart's words (in Sen, 1965, p. 218), 'the greatest blessing any nation can enjoy is a uniformity of opinion on every point which concerns public affairs and the administration of them.'[18] That explains the virtually catholic emphasis on the merits of discipline among Steuart's contemporaries as well.

But this is only one strain of 18th-century thought. Smith, on the other hand, linked freedom with the right to own property and with the individual rights and restrictions of the private person:

> Every man, as long as he does not violate the laws of justice, is left perfectly free to pursue his own interest his own way, and to bring both his industry

[17] Of course today no one would think the man who has been so molded and manipulated that he always wants what his superior or governor wants is free. In the eighteenth century this concept of freedom was juxtaposed with slavery and the molding was viewed as necessary to maintain political stability.

[18] I was, unfortunately, unable to locate this quotation in the 1805 edition of Steuart's collected works.

© Scottish Economic Society 1996

> and capital into competition with those of any other man, or order of men (Smith, 1976a, p. 687).

With this view Smith was following the Physiocrats, for whom

> freedom implies the abolition of privileges and regulations in all markets. Free competition must rule in the labour market as well as in domestic and foreign trade. ... For the Physiocrats freedom meant universal competition, and was regarded as the basis for the increase in private and public wealth (Vaggi, 1987, p. 873).

These two competing concepts of liberty in the 17th and 18th centuries serve to explain not only Steuart's views but also Smith's equivocal attitude towards the system of natural liberty. Needless to say, Smith and Steuart lived in an age of transition, both absorbing parts of the new and old traditions in differing degrees. Probably because of his absence from Britain, Steuart's work, so aptly characterized by Roll (1938, p. 127), 'breathes little of that air of unbridled self-interest and freedom of trade that was common [in England] at the time.'

David Hume, Smith and Steuart's mutual friend, also mentioned in his economic writings that the republic of Sparta was 'justly esteemed by every one' (1955, p. 8) and acknowledged the necessity of public indoctrination, the *amor patriae*.[19] But his further insights into the nature of Sparta are particularly enlightening for an understanding of Steuart's position. Hume argued that Sparta had to be a martial state by virtue of its geography: all of its neighbours were 'continually in arms' (*ibid.*, p. 9). He argued further that the legislator should never go against the nature of the people:

> the best policy [is] to comply with the common bent of mankind, and give it all the improvements of which it is susceptible' (*ibid.*, p. 10).

In his *Principles* Steuart repeatedly emphasized the statesman's need to attend closely to the spirit of the people. Thus, by a 'perfect plan' it would not be unreasonable to suggest that Steuart meant two things: first, that the spirit of the citizens of Sparta was most perfectly reflected in the 'plan of the economy,' and second, that the subordination of the Spartan citizens to their state was unparalleled in history. An interpretation that avoids an association with dictatorial rule gains further credence if we keep two things in mind. Hume read the draft of the *Principles* and was 'exceedingly pleased with it' (Skinner's introduction to Steuart, 1966, p. xlv) and the American edition was admired by both Alexander Hamilton and Benjamin Franklin (King, 1988,p. 31).[20]

[19] And thus Anderson and Tollison's (1984, p. 464) assertion that 'Steuart's enthusiasm for totalitarian economic planning was apparently unique among English as well as continental thinkers contemporary with Smith' falls, both in the sense that Steuart allegedly propagated a totalitarian regime and that Steuart alone admired Lycurgus's Sparta.

[20] A brief excerpt from Hamilton's work that is reminiscent of Steuart's themes has been reprinted in Prestowitz, Morse and Tonelson's *Powernomics* (1991, pp. 129–46). See also Spiegel (1987). Both Alexander Hamilton and Abraham Lincoln were for the protection of infant industries (Kuttner, 1991, p. 31). Even Smith scholars forget that Smith's system of natural liberty was flexible enough to embrace the protection of infant industries under certain conditions (Smith, 1976a, p. 37 and 463ff.).

© Scottish Economic Society 1996

The real nature of government

Steuart's statesman deserves closer scrutiny. The label is, Steuart (1966, p. 16, I, introduction) tells us, 'a general term to signify the legislature and supreme power.' Steuart (1966, p. 708, V, vii) acknowledged that the term is often used as an idealized abstraction: 'In my inquiries, I have constantly in my eye, how man *may* be governed, and never how *he is* governed.'[21] Eltis (1986, p. 68) makes a point with his remark that it is 'precisely because Smith described how man *is* governed, while Steuart merely sought to show how he *may* be governed that his analysis carried extra conviction to his contemporaries.' Chamley (1963, p. 79) has raised the question whether it is not strange for Steuart to have relied on an abstraction in a work permeated with historical relativism.

Steuart always supposed the statesman's 'inclinations to be virtuous and benevolent' (1966, p. 333, II, xxvii) and described him this way:

> I suppose him to be constantly awake, attentive to his employment, able and incorrupted, tender in his love for the society he governs, impartially just in his indulgence for every class of inhabitants, and disregardful of the interest of individuals, when that regard is inconsistent with the general welfare (1967a, vol. 1, p. 200, I, xxi).

The government's purpose is to channel positive energy so that the general welfare is maximized. Similarly, Steuart's task was

> to investigate how a statesman may turn the circumstances which have produced this new plan of oeconomy to the best advantage for mankind, ... to examine the consequences of what we feel and see daily passing, and to point out how far the bad may be avoided, and the good turned to the best advantage (Steuart, 1966, pp. 75–76, I, xii).[22]

Hence, the government is clearly benign in Steuart's scheme of things, rather like a paternalistic overseer. That in fact is precisely the analogy Steuart (1966, p. 15, I, introduction) drew upon: 'The whole economy must be directed by the head, who is both lord and steward of the family,' and furthermore, 'oeconomy in general [as opposed to political oeconomy] is the art of providing for all the wants of a family, with prudence and frugality' (Steuart, 1966, p. 15, I, introduction). The comparison of family to state is applied liberally throughout the book, with Steuart (1966, p. 68, I, xi) admitting that he 'often recur[s] to private oeconomics for clearing up my ideas concerning the political.'

Steuart cautioned furthermore that the statesman is not omnipotent because he

[21] Plato discussed politics in terms of an imaginary ideal state, Steuart in terms of an imaginary ideal statesman.

[22] Consider also this comment of Steuart (1966, p. 122, I, xix): 'In treating every question of political oeconomy, I constantly suppose a statesman at the head of it, systematically conducting every part of it, so as to prevent the vicissitudes of manners, and innovations, by their natural and immediate effects or consequences, from hurting any interest within the commonwealth.'

© Scottish Economic Society 1996

is constrained by a consideration of the spirit of the people, their morals, government, and manners, which form the basis for the laws, customs, and form of government of the country. By the same token, state management cannot be characterized as anti-democratic. To the contrary, Steuart believed the modern economy:

> is the most effectual bridle ever ... invented against the folly of despotism; so the wisdom of so great a power never shines with greater lustre, than when we see it exerted in planning and establishing this oeconomy as a bridle against the wanton exercise of itself in succeeding generations (1966, p. 279, II, xxii).

Steuart guarantees, furthermore, that the statesman:

> is neither master to establish what oeconomy he pleases, or, in the exercise of his sublime authority, to overturn at will the established laws of it, let him be the most despotic monarch upon earth (Steuart, 1966, p. 16, I, introduction).[23]

But it cannot be denied that Steuart was an advocate of a strong, centralized government and a principle of extensive public planning far surpassing Smith's or later Anglo-Saxon notions of state power. Why this was so can be explained by the German influence.

Although Steuart followed Montesquieu in some ways, his theory of the state was not a French one. In that era three trends in economics were prominent in France: the Encyclopedists' emphasis on reason, the physiocrats' focus on reform and a minimal role for the state, and the egalitarians' (e.g., Rousseau's) focus on building utopias. All three trends glorified the individual over the state and advocated the free reign of natural forces to bring the best outcome—the basis for Smith's economics. Steuart in contrast believed citizens should subordinate themselves to an enlightened government that assumed a protective, rather paternal, function. As Steuart (1966, p. 20, I, i) remarked, 'the end of a *voluntary* subordination to authority is with a view to promote the general good.'

This view of the promotion of the general welfare was common in Steuart's age and is the philosophical underpinning of the *Principles*. European societies emerged from feudalism with an extremely sharp distinction between ruler and ruled; the corresponding values of the state encouraged individuals to defer to the government. In contrast, the United States, without a feudal history, based

[23] The exalted role of the statesman in Steuart's works does not mean the statesman is altogether eliminated in Smith's works. Smith discusses three types of statesmen, only two of which play a negative role: politicians and 'men of systems.' His legislator, on the other hand, is depicted as 'the greatest and noblest of all characters, that of the reformer and legislator of a great state' (Smith, 1976b, p. 23) and thus seems not far removed from Steuart's concept of the statesman. See West (1976, especially pp. 627–33) and Winch (1978, especially pp. 159–60, 171–73, and 181). Recall also that for Smith intervention in the economy meant intervention on behalf of the wealthy and powerful. Only by Mill's time was governmental intervention on behalf of the poor conceivable.

© Scottish Economic Society 1996

its government on the idea of widely dispersed powers and checks and balances, which in turn generated a corresponding ethic sceptical of government, a belief in private over public interests, and rugged individualism. Even today the deference to the government displayed by the citizens of these same European states—most of which are now democratic, free-market societies—far surpasses that which Americans would consider healthy. The same goes for their system of developing economic policies. To mention one example, German *Wirkschaftspolitik*—the policy measures taken by the state concerning the economic process, structure, and order—is unknown in today's Anglo-Saxon world, which has no such tradition and, being anti-planned economies, even shuns a systematic development of policy (i.e., a national economic strategy).

Steuart's theory of the role of government, although flippantly dismissed as mercantilist leanings by many a Steuart student, seems to have been borrowed wholesale from the German cameralists' theory of the state. Steuart probably absorbed their doctrines while in Tübingen.[24] Cameralism was a specific strain of mercantilism peculiar to the Austro-German principalities of the 17th and 18th centuries. In Steuart's lifetime a mature cameralism was just evolving into *Staatswissenschaft*, a translation of the term *res politicae* made popular by the cameralists.[25] The object of *Kameralismus* and *Staatswissenschaft* was the state, its goals, and its tasks. Austro-German *Kameralismus* emphasized the paternalistic nature of the government's policy plan, whereas the French and English strains of mercantilism focused on the aggrandizement of the wealth of the state as an end in itself. For the cameralists the state was seen as an organic whole whose various economic branches—*Landwirtschaft* (agriculture), *Gewebe* (industry), *Manufaktur* (manufactures), *Verkehr* (transport), and *Handel* (trade)—needed coordination. Similarly, political economy for Steuart was, as the subtitle of his *Principles* indicates, 'the science of domestic policy in free nations in which are particularly considered population, agriculture, trade, industry, money, coin, interest, circulation, banks, exchange, public credit, and taxes.'

Steuart and the cameralists were answering the much debated question of the ancient Greeks: Which form of policy and government is the best (in terms of raising the general welfare)? His welfare concept closely matches that of Johann Heinrich Gottlob von Justi (1717–1771), Steuart's contemporary and 'the great systematizer and perfecter of cameralism' (Engelhardt, 1981,

[24] 'It would seem almost impossible that Steuart would not come in contact with cameralist writings during this period [i.e., during his stay in Tübingen], especially since Chairs of Cameralistics had been established as early as 1727' (Skinner's introduction to Steuart, 1966, p. xl, n. 94).

[25] Today's German Ph.D. in economics, the *Doctor rerum politicarum*, is a hangover from this tradition, as is the *venia legendi* for *Wirtschaftliche Staatswissenschaften*, the terminal degree still conferred by some Germany economic faculties on those who complete the requirements in economics for the *Habilitation*, subsequently giving them the right to take up a professorial position at the university. A later outgrowth of this tradition was the journal *Zeitschrift für die gesamte Staatswissenschaft*, founded in 1844, and since 1986 unfortunately renamed the *Journal of Institutional and Theoretical Economics* (*JITE*).

© Scottish Economic Society 1996

p. 47).[26] To the cameralists the state had a *Polizey-Tätigkeit*, a policy-making function whose goal was to bring about the happiness of all. Justi's message has been summed up this way (Tautscher, 1959, p. 760): 'Der Staat wird der oberste Lenker aller sozialwirtschaftlichen Vorgänge und bringt dadurch den Wohlstand und die "Glückseligkeit" aller hervor' (The state is the foremost controller of all social-economic processes and thus brings about the welfare and 'happiness' of all). For both Justi and Steuart the foremost purpose of the state was the guarantee of the *Glückseligkeit* or happiness of all its subjects.[27] While cameralism before Justi had focused on a certain region or a *Fürst*'s (prince's) territory, Justi turned his attention to the nature of the state, basing *Staatswissenschaft* on one principle: the subjects' happiness. Thus, the mission of Justi's *Staatswirthschaft* ('National Economy') was to explain how a ruler should govern in order to maintain a flourishing population and assure the happiness of his subjects.

The fact that people are understood as being motivated by *Glückseligkeit* in Steuart's *Principles* has two ramifications. First, it is responsible for a *Willensvereinigung*, a voluntary union of will providing the state with the power to intervene on its subjects' behalf. Second, the striving for happiness is the force or impulse behind economic activity. That is, Steuart sees in self-interest and a desire for material well-being the sources of industry. This assumption that men work because they aspire to a certain level of consumption—dubbed the 'aspiration effect' by Eagley (1961)—became the basis of Steuart's aggregate demand theory.

What must the state provide for the happiness of its subjects? According to Justi, who once said 'the happiness of the people ... is my only wish' (in Engelhardt, 1981, p. 67), the subjects can only be happy given the following four conditions: internal peace and peace with bordering countries, freedom (meaning no subjects are slaves), a functioning judicial system, and the opportunity to subsist or do better in a flourishing economy (Engelhardt, 1961). If the state, representing the union of wills and strength of its citizens, is to fulfill these conditions, it obviously must be powerful. The first three conditions are generally taken for granted by Steuart; setting forth the best plan to guarantee the last condition is the goal and substance of his *Principles*. Thus, with Steuart we enter an age of economics.

The policy goals of cameralism were directed to the rebuilding of a country ravaged by the Thirty Years' War. Hence, most cameralist policy arguments cen-

[26] Schumpeter (1954, p. 176) reached a similar conclusion: 'Steuart's work did not ride, like Smith's on the wave of a single and simple policy that was rapidly conquering public opinion. On the contrary, he grouped all that really interests the public around the old-fashioned figure of an imaginary patriot statesman who in infinite wisdom watches the economic process, ready to interfere in the national interest—a conception that recalls Justi's and was quite out of contact with England's humor.' See also Skinner's commentary in his preface to Steuart 1966 (p. xl, n. 94).

[27] Justi (1758, pp. 35–36) upheld the view that 'the first, general principle in all economic and cameral sciences' is 'that the whole governmental business of a state is arranged in such a way that the happiness [of its subjects] must be supported.' ('Hieraus folget der erste und allgemeine Grundsatz aller öconomischen und Cameralwissenschaften, nämlich: daß alle Regierungsgeschäfte eines Staats solchergestalt eingerichtet werden müssen, daß dadurch die Glückseligkeit desselben befördert werde.')

© Scottish Economic Society 1996

tred around such practical measures as 'an active population policy, the establishment of state manufactures and banks, the extension of infrastructure (canals, bridges, harbours and roads) and the promotion of modernization' (Recktenwald, 1987, p. 314). But the major policy goal of the state under cameralism was the full employment of its citizens. In the *Principles* Steuart adopted the cameralistic program with its focus on the welfare of its subjects. As a result, state intervention and policy considerations for Steuart—to apply Schumpeter's assessment of Justi's position (1954, p. 172) to Steuart—is closer to 'a laissez-faire plus watchfulness, ... [rather] like laissez-faire with the nonsense left out.'

Although the Anglo-Saxon welfare notion differs radically from this, the 18th-century German welfare concept was not discarded in America but modified to fit individualism: all Americans received the right by constitutional law to the 'pursuit of happiness.' This, of course, turns Steuart's cameralist concept on its head since the state takes a subordinate role to the individual and the welfare and happiness of its citizens; the government's role is even intentionally limited to protect infringements on individual freedoms. While this helps to explain how Anderson and Tollison reached the conclusion that Steuart's economics was the converse of Smith's, *Glückseligkeit* does not represent the pure converse but the roots of the modern concept of welfare economics. Because common welfare depended upon a conformity of wills between governed and governing, the question, Is the essence of welfare to be decided by the people and communicated to the government or vice versa? arises. This in turn opened up the question of representative government, not dealt with by the cameralists or Steuart and which, of course, also received a different answer on American soil.

The kinship with Keynes

There are several other plausible reasons for Steuart's giving government such a broad reign to intervene in the economy—reasons that give Steuart the distinction of being a precursor of Keynes. Both Smith and Steuart were taken in by 'the whole magic of a well-ordered society,' to use a common expression of the Physiocrats (cf. Hébert, 1987, p. 201). But while Smith's assumption that the existing social order did not correspond to the natural order led him to argue, following French physiocratic doctrine, that economic forces should be left to their natural course, Steuart was convinced that only the government could achieve order. His commitment to intervention in fact derives from a *rejection* of the natural law doctrines so popular in eighteenth-century England and France. In Steuart's opinion:

> fundamental, that is invariable laws, can never subsist among men, the most variable thing we know: the only fundamental law, *salus populi*, must ever be relative, like every other thing.[28] (1966, p. 21, I, i)

[28] Johnson (1960, p. 211) I think rightly argues: 'His reading of Montesquieu and his own experiences on the Continent convinced him that no principles of government and no economic policy had universal application. Foremost in his political theory stands an emphasis upon adaptation, a regard for the 'spirit of a people'. ...' In this respect Steuart shares much in common with a fellow Scotsman and contemporary, David Hume.

© Scottish Economic Society 1996

Like his mercantilist forerunners and Keynesian predecessors, Steuart viewed the economy as inherently unstable, lacking natural forces or self-adjusting mechanisms that guarantee economic order. Due to the interdependence of the economy, order in his view was too complicated to be a natural phenomenon. Hence the need for a controlling authority, 'the statesman,' who was to act as a balancing force on the economy.

Another reason for giving government such wide leeway was his understanding of equilibrium. Equilibrium for Steuart is not the balance of opposing forces, the equation of the supply with the demand of goods, a state where things are at rest, or a central tendency. It is the economic order best satisfying human needs. To Steuart a balanced economy was good; disequilibrium was equated with stagnation and decline. Steuart was convinced that the system very often moves away from a balanced position—the reason for putting adjustment in the statesman's hands.

In Steuart's view the chief task of the government is 'to provide food, other necessaries and employment, not only for those who actually exist, but also for those who are to be brought into existence' (1967a, vol. 1, p. 200, I, xxi), making full employment and the full utilization of all productive powers of the economy priority policy goals.[29] As Steuart saw it, especially the labour market was unstable; thus the statesman was to effect 'a vibration of the balance between work and demand.' The existence of an excess supply of labour could in his view be alleviated by public works (which would in turn extend and strengthen the economic infrastructure). Because Steuart's concept of happiness assumed at least a subsistence living level, a distribution problem naturally emerged whose solution Steuart saw in taxation.[30]

The family resemblance to Keynes nonetheless has its limits. With both Steuart and Keynes there is a shift to the demand side and a priority support for eliminating unemployment as a social objective. But Steuart followed the cameralists in forming policy recommendations that

> emphasized the paternalist character of the government's centralized fiscal policy (not, as is sometimes mistakenly thought, a Keynesian short-run instrument but rather a regulator for development which was to serve the general happiness of the subjects ...) (Recktenwald, 1987, p. 313).

That is the reason why we today find Steuart's theory of the state, swallowed whole, unrealistic.

Supranational politics, international economics

Steuart's statesman is neither nationalist nor restricted by state boundaries. Political economy, the job of the statesman, is in Steuart's view a 'great art'

[29] Compare the Keynes of the *General Theory* (1973, p. 372): 'The outstanding faults of the economic society in which we live are its failures to provide for full employment and its arbitrary and inequitable distribution of wealth and incomes.'

[30] 'Steuart was of course far ahead of his time in his eulogies of high taxation, and in the emphasis he placed on the public sector' (Eltis, 1986, p. 62).

© Scottish Economic Society 1996

whose mission is 'first to adapt the different operations of it to the spirit, manners, habits, and customs of the people; and afterwards to model these circumstances so, as to be able to introduce a set of new and more useful institutions' (1966, p. 16, I, introduction). For Steuart there is no pat policy solution to economic problems since each country and each economic stage demands the development of different rules and policies adapted to its distinct circumstances:

> If one considers the variety which is found in different countries, in the distribution of property, subordination of classes, genius of people, proceeding from the variety of forms of government, laws, climate, and manners, one may conclude, that the political oeconomy in each must necessarily be different ... (Steuart, 1966, p. 17, I, introduction).

Since Steuart always worked within a European framework, it was an easy step to extend state management and the welfare concept to the supranational level. Steuart took this step with the advocacy of joint economic coordination among states aimed at guaranteeing free trade and the 'common good' for an entire geographical region:

> Nothing, I imagine, but an universal monarch, governed by the same laws, and administered according to one plan well concerted, can be compatible with an universally open trade. While there are different states, there must be separate interests; and when no one statesman is found at the head of these interests, there can be no such thing as a common good, every interest must be considered separately (Steuart, 1966, p. 365, II, xxix).

With this call for a Europe united by free trade Steuart was perhaps the first to suggest the creation of a European Community, although in so doing Steuart was reiterating Montesquieu's defence of free trade on a European scale.[31] Clearly ahead of its time in recognizing the profits that could be won from cooperation, Steuart's discussion of a coordination problem at the European level takes on a new, practical dimension with the opening of an integrated European Common Market in 1992.[32]

V THE STEUART LEGACY

In the foregoing we have tried to explode the myths surrounding Steuart's theory of the state. In what remains we will illuminate the consequences for this mistaken understanding of Steuart's statesman and welfare concept. History's

[31] Skinner (1963, p. 447), quoting Montesquieu, also notes that 'Steuart agreed with Montesquieu that trade would tend to make all the world "one single state of which all the societies on earth are members" ...'

[32] Steuart's focus on coordination both on a national and supranational level exposes Johnson's charge of economic nationalism as erroneous. Johnson (1965, p. 214) argues: 'In spite of his warning that the economist must be a citizen of the world, and in spite of his exile, Steuart was a staunch nationalist; the principles which he sought were those which underlay "domestic policy." ' Besides the reference to 'domestic policy,' no other evidence of nationalism is provided.

© Scottish Economic Society 1996

judgment, we will see, has been too harsh on Steuart, virtually eclipsing his achievements.

One sure measure of the depth of an economist's thought is the test of time: theoretical tools or policy recommendations sound enough to survive generations evidence economic acumen. Steuart's achievements—many of which form the basis of modern economic theory and policy—have been overlooked or underrated due to the unwarranted commotion raised over his statesman. A full treatment of Steuart's theoretical contributions oversteps the boundaries of this paper, but an overview will provide us with a glimpse of the lost legacy.

Steuart's work went unappreciated in Britain after Smith's *Wealth of Nations* appeared, prompting Steuart (in Skinner's introduction to Steuart, 1966, p. lv) to remark, shortly before his death, that a book on the life of his dog would have been more 'relished by the public.' But it is Steuart's theory of the state, along with his historical bent, that made a warm reception in the German-speaking world possible. With hindsight it must be considered an unfortunate quirk of history that the reaction to Steuart's statesman, both from Steuart's countrymen and the Anglo-Saxon world in general, was so polemical in nature (even going so far as to endeavor to discredit him), while the judgment passed in the German-speaking world was far more balanced and reliable. Several German economists—Ingram (1926) and Sen (1957, p. 187) mention Hufeland, Hasbach, Herrenschwand, Rehberg, and Roscher—extolled Steuart's work *over Smith's* for its historical relativism, emphasis on the spirit of the people, and reservations about making sweeping generalizations. Roscher (1874, p. 563), in addition, ranked Steuart as 'doubtless a great economist' who did not deserve to be overshadowed by Smith, whose accomplishments in his view were overrated in both the Germany and England of his day (Roscher, 1867).[33] As early as 1786 Herrenschwand (1980, p. 306) judged Steuart as 'the most profound of all the English writers in political economy.' Like many of his contemporaries, Rehberg (1829, p. 302) argued that Steuart 'was not as well known in Germany as he deserved to be,' adding that Steuart should be praised for 'a talent for abstraction and scientific work seldom encountered.' Hasbach (1891, pp. 369–81) dedicated a section encompassing 13 pages to Steuart in his book on Adam Smith. He credited Steuart for separating economics and ethics and concluded that Steuart was 'the greatest economist of the 18th century' and one of the greatest of all times (p. 381). Zuckerkandl (1889, pp. 148–52) and Berens (1868, pp. 33–38) each dedicated somewhat more than five pages to Steuart's contributions, with Berens (1868, p. 38) pointing out how Steuart clearly anticipated von Thünen's key ideas. Commending Steuart's work for being 'rich in observations and sagacious comments,' Weitzel (1832, part 1,

[33] See also Hasek (1925) for the reception of Smith's *Wealth of Nations* in Germany. While the University of Göttingen led in the transmission of Smith's ideas, at Tübingen the discipline of cameralism was just becoming established. There the first chair in cameralism was erected in 1817, although courses in cameralism were offered from 1717 on—first by the botanist Christian Reuse and then by Friedrich Fulda, a member of the law faculty (see Conrad, 1960). Both Smith and Steuart are mentioned in Fulda's formal pleas of 1805 for a chair in cameralism at Tübingen. Steuart probably associated with professors representing a wide range of disciplines, including law, while living in Tübingen.

© Scottish Economic Society 1996

p. 256) put both Smith and Steuart together in one section of his history of *Staatswissenschaft* and allotted them equal space. In his day a distinguished historian of economic thought, Kautz (1860, p. 286) praised Steuart for having discovered 'numerous truths in the science [of political economy]' and predicted Steuart would have 'an important and lasting influence' on economic theory.

Beyond this, Steuart's work would go on to make a sharp imprint on two famous German personalities. Hegel (1770–1831) drew on Steuart's stadial model of economic development to show how commerce had produced a completely new society.[34] Friedrich List (1789–1846), like Steuart, remained almost unknown in the Anglo-Saxon world but exerted considerable influence in the German-speaking countries. Steeped in Alexander Hamilton's writings, List became a proponent of the protection of infant industry while living in the United States. Viewing Smith's system of natural liberty as an erroneous generalization of Britain's situation to the rest of the world, List echoed Steuart in his conviction that countries should be equally developed before free trade can function to everyone's mutual advantage. Finally, to close this sampling of German opinion, Schumpeter (1954, p. 176) appraised Steuart's theories of population, prices, money, and taxation as going 'much below the smooth surface on which A. Smith sailed his course.'

Even in Britain and North America Steuart's ideas found a following, when not acknowledged. Malthus would go on to develop Steuart's population theory and take up his concept of effective demand. During the zenith of the Keynesian revolution Steuart enjoyed a revival: Steuart's monetary and employment theory have won the praise of, among others, Hutchison, Schumpeter, and Vickers for its anticipation of Keynesian theory. Steuart's views on public debt were commended by Stettner. Viner (1937, p. 10) noted that the term favourable balance of trade 'seems first to have been used in 1767 by Sir James Steuart.' Steuart's policy statement on grain written in 1759 in Tübingen—a proposal for the 'Common Markets of England'—has, as Eltis remarks, 'become the extremely controversial policy of grain of the European Common Market.'[35] Eltis adds that the Cambridge Economic Policy Group's analysis of British industries in the 1970s echoes Steuart's solutions to averting both a decline in international competitiveness and high unemployment. Kuttner's reasoning in *The End of Laissez Faire* (1991) reverberates with Steuart-type arguments for state intervention, although Kuttner gives Keynes, List, and Karl Polanyi as the sources of his inspiration.

The real legacy, however, has been the realization of Steuart's vision of an international basis for economic cooperation and coordination of policy matters. Steuart placed Europe in an age of economics; questions of peace, freedom, and legal matters essentially paled into the background. Today, with the fall of

[34] For more on the relationship between Steuart and Hegel's ideas see Chamley (1963).

[35] Steuart proposed that the government buy up all the grain farmers produce at the minimum price and sell as much as possible at a maximum price. Unsold grain was to be stored in state granaries. Today's European policy is, by the way, far more bureaucratic than Steuart ever could have envisioned (e.g., each animal is tagged and controlled).

© Scottish Economic Society 1996

communism, the breakup of the Eastern Bloc and the USSR, and the emergence of an integrated European Common Market in 1992, we are entering upon a new age of economics with international dimensions that—unbeknownst to most students of economics—fulfills Steuart's vision.

The recognition of the immediacy of a process by which countries adjust their monetary and fiscal policies and exchange rates in a mutually beneficial manner actually predates the fall of communism. The Group of Seven countries (G-7) has been making progress in this area since the breakdown of the Bretton Woods system. Coordination has been exemplified most recently by the Bonn Summit of 1978 (where Bonn agreed to a policy of macroeconomic reflation, Washington to raise oil taxes to conserve energy and reduce dependence on OPEC) and the Plaza Accord of 1985 (where agreements were made to manage fluctuations in the dollar). In addition, the G-7 are considering coordinating their central banks, while the G-3 (US, Japan, and Germany) have informally agreed to exchange-rate zones. All this represents an informal arrangement that may well develop into a more concrete formal system in the future.[36] To be sure, the recognition that noncooperative behaviour on a world scale is inefficient has become common knowledge.

The basis for economic coordination and cooperation has by no means been focused solely on the international level: it is quite consistent with and even presupposes a domestic economic strategy. Thus, Steuart's legacy, Anderson and Tollison notwithstanding, is also manifested in a limited but growing understanding—even in the Anglo-Saxon world—that the development of a national economic strategy can be compatible with a free-market economy and bring great benefits.[37]

It should be clear, then, that Steuart and Smith's theories of the state (extricated from the particular version of freedom) represent two sides of the same coin, both answering the question, What kind of capitalism do we need? In many ways their answers represent two extreme endpoints of the spectrum of capitalism, Steuart's being a heavy mix of state to private activity and Smith's the converse. Steuart's primary social objective was employment and stabilization while Smith focused on individual liberties; Steuart's regulatory framework hovers in the foreground while Smith's recedes into the background. In spite of their incongruities, Steuart and Smith's theories of the state are the legacy of Western market economies. Even the 20th-century Anglo-Saxon world has, historically, oscillated between these extremes as social objectives and views on the proper role of the state were modified. Steuart's statesman reached its apotheosis when the Keynesian view that government has an obligation to its constituency to maintain levels of aggregate demand sufficient to guarantee universal employment became legally anchored in national statutes (for instance, in the US with the passing of the Employment Act of 1946 that committed government to support employment policies). Economists must also

[36] See Currie and Levine (1991) and de Grauwe (1987) for a discussion of these developments.

[37] See, for instance, the contributions in Prestowitz, Morse, and Tonelson's *Powernomics* (1991).

© Scottish Economic Society 1996

admit that the choice of implementing active fiscal or monetary policy or placing a faith in Say's law (automatic adjustment) not only distinguishes Steuart's economics from Smith's, but remains *the chasm* dividing the contemporary views on theory and policy.[38]

ACKNOWLEDGEMENTS

This contribution is dedicated to the foremost Steuart scholar, Andrew S. Skinner, without whose generous assistance it could not have been completed. A special thanks also goes to Joachim Starbatty, whose comments on the German, abridged version helped to tighten the argument.

REFERENCES

AKHTAR, M. A. (1978). Steuart on growth. *Scottish Journal of Political Economy*, 25, 1, pp. 57–74. Rpt. in Blaug (1991), pp. 207–24.

AKHTAR, M. A. (1979). An analytical outline of Sir James Steuart's macroeconomic model. *Oxford Economic Papers*, 31, 2, pp. 283–302. Rpt. in Blaug (1991), pp. 225–44.

ANDERSON, G. M. and TOLLINSON, R. D. (1984). Sir James Steuart as the apotheosis of mercantilism. *Southern Economic Journal*, 51, October, pp. 450–68.

BERENS, E. (1868). *Versuch einer kritischen Dogmengeschichte der Grundrente*. Leipzig: K. F. Köhler.

BLAUG, M. (ed.) (1991). *Pioneers in Economics*, 11, D. Hume (1711–1776) and J. Stewart (1712–1780). Aldershot, Hants & Brookfield, Vermont: Elgar.

CHAMLEY, P. (1962). Sir James Steuart, inspirateur de Lord Keynes? *Revue d'économie politique*, May–June, pp. 303–13.

CHAMLEY, P. (1963). *Economie politique et philosophie chez Steuart et Hegel*. Paris: Dalloz.

CONRAD, E. (1960). Die Lehrstühle der Universität Tübingen und ihre Inhaber (1477–1927). Thesis leading to a teaching degree, Tübingen University.

CURRIE, D. and LEVINE, P. (1991). The international co-ordination of macroeconomic policy. In *Companion to Contemporary Economic Thought*, D. Greenaway, M. Bleaney, and I. M. T. Stewart (eds.), pp. 482–506.

DAVIE, G. E. (1967). Anglophobe and Anglophil. *Scottish Journal of Political Economy*, 14, November, pp. 291–302.

DE GRAUWE, P. (1987). International monetary policy. In *The New Palgrave: A Dictionary of Economics*, 1, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

DEVLETOGLOU, N. E. (1963). Montesquieu and the *Wealth of Nations*. *Canadian Journal of Economics and Political Science*, 24, 1, February, pp. 1–25. Rpt. in *Pre-Classical Economists*, II, pp. 37–61. M. Blaug (ed.) 7, Pioneers in economics. Elgar: Aldershot & Brookfield (1991).

EAGLY, R. V. (1961). Sir J. Steuart and the aspiration effect. *Economica* (NS) 18, February. Rpt. in Blaug (1991), pp. 118–26.

ELTIS, W. (1986). Sir James Steuart's corporate state. In *Ideas in Economics* (The Proceedings of the 1985 Meeting of Section F of the British Association), R. D. C. Black (ed.). London: Macmillan, pp. 43–73.

ELTIS, W. (1987). Steuart, Sir James. *The New Palgrave: A Dictionary of Economics*, 4, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

ENGELHARDT, U. (1981). Zum Begriff der Glückseligkeit in der kameralistischen

[38] It only seems fair to me that Steuart be placed on an equal footing with Smith in the history of economic thought for both his contribution to economic systems and their coordination and for the first development of principles of political economy.

© Scottish Economic Society 1996

Staatslehre des 18. Jahrhunderts. *Zeitschrift für historische Forschung*, 8, 1, pp. 37–79.

FEILBOGEN, S. (1889). James Steuart und Adam Smith. *Zeitschrift für die gesamte Staatswissenschaft*, 45, pp. 218–60.

FULDA, F. K. VON. (1805). *Ueber das Kameralstudium in Wirtemberg. In vier Briefen*. n.p.

HASBACH, W. (1891). *Untersuchungen über Adam Smith und die Entwicklung der Politischen Ökonomie*. Leipzig: Duncker & Humblot.

HASEK, W. C. (1925). *The Introduction of Adam Smith's Doctrines into Germany*. New York: Longmans, Green & Co. and London: P. S. King & Son, Ltd.

HÉBERT, R. F. (1987). In search of economic order: French predecessors of Adam Smith. In *Pre-Classical Economic Thought from the Greeks to the Scottish Enlightenment*, S. T. Lowry (ed.). Boston, Dordrecht, Lancaster: Kluwer Academic Publishers, pp. 185–220.

HERRENSCHWAND, J. (1980) [1786]. *De l'économie politique moderne. Discours fondamental sur la population*. München: Kraus Reprint.

HUME, D. (1955). *David Hume: Writings on Economics*. Edited and introduced by E. Rotwein. Edinburgh, London, Melbourne, Cape Town, etc.: Thomas Nelson and Sons.

HUTCHINSON, T. W. (1988). *Before Adam Smith: The Emergence of Political Economy, 1662–1776*. Oxford and Cambridge, MA: Basil Blackwell, ch. 19.

INGRAM, J. K. (1926). Steuart, Sir James. *Palgrave's Dictionary of Political Economy*, 3. London: Macmillan.

IRWIN, W. R. (1962). Eine schottische Adelsfamilie in Tübingen. *Tübinger Blätter*, pp. 42–43.

JOHNSON, E. A. J. (1960) [1937]. *Predecessors of Adam Smith: The Growth of British Economic Thought*. NY: Prentice-Hall, ch. 11: Steuart, the political economist.

JUSTI, J. H. G. VON. (1758). *Staatswirthschaft oder systematische Abhandlung aller oekonomischen und Cameralwissenschaften*, 2 vols. Leipzig: Bernhard Christoph Breitkopf.

KAUTZ, J. (1860). *Theorie und Geschichte der National-Oekonomik*. Part II: Die National-Oekonomik als Wissenschaft. Wien: Carl Gerold's Sohn.

KEYNES, J. M. (1973). *The General Theory of Employment, Interest and Money*, 7: The Collected Works of John Maynard Keynes. London & Basingstoke: Macmillan for the Royal Economic Society.

KHALIL, E. L. (1987). Sir James Steuart vs. Professor James Buchanan: critical notes on modern public choice. *Review of Social Economy*, 45, 2, October, pp. 113–32.

KING, J. E. (1988). *Economic Exiles*. Basingstoke, London: Macmillan.

KOBAYAYSHI, N. (1967). *Sir James Steuart, Adam Smith and Friedrich List*. Tokyo: The Science Council of Japan.

KUTTNER, R. (1991). *The End of Laissez-Faire: National Purpose and the Global Economy After the Cold War*. New York: Alfred A. Knopf.

LETWIN, W. (1963). *The Origins of Scientific Economics: English Economic Thought 1660–1776*. London: Methuen.

MEEK, R. L. (1958). The economics of control prefigured by Sir James Steuart. *Science and Society*, 22, 4, pp. 289–305. Rpt. in Blaug (1991), pp. 101–17.

PRESTOWITZ, C. V. Jr., MORSE, R. A. and TONELSON, A. (1991). *Powernomics: Economics and Strategy After the Cold War*. Lanham, MD: Madison Books for the Economic Strategy Institute.

RASHID, S. (1986). Smith, Steuart and mercantilism. *Southern Economic Journal*, 52, 3, January, pp. 843–52, with Reply from Anderson and Tollison, pp. 853–55.

RECKTENWALD, H. C. (1987). Cameralism. In *The New Palgrave: A Dictionary of Economics*, 1, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

REHBERG, A. W. (1929). Ueber Steuart's principles of political oeconomy und Smith on the wealth of nations. In his *Politisch-historische kleine Schriften*. Hannover: Hahn'sche Hofbuchhandlung, pp. 299–309.

RICHTER, M. (1977). *The Political Theory of Montesquieu*. Cambridge, London, New York, Melbourne: Cambridge University Press.

ROLL, E. (1938). *A History of Economic Thought*, rev. and enlarged. London: Faber and Faber Ltd.

© Scottish Economic Society 1996

ROSCHER, W. (1874). *Geschichte der National-Oekonomik in Deutschland*. München: R. Oldenbourg.

ROSCHER, W. (1867). *Die Ein- und Durchführung des Adam Smith'schen Systems in Deutschland*. Dresden: Berichte der kön. Sächs. Gesselschaft der Wissenschaft.

RYAN, A. (1987). Liberty. In *The New Palgrave: A Dictionary of Economics*, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

SCHUMPETER, J. A. (1954). *History of Economic Analysis*. New York: Oxford University Press.

SEN, S. R. (1947). Sir James Steuart's general theory of employment, interest and money. *Economica*, 14, February (NF), pp. 19–36.

SEN, S. R. (1957). *The Economics of Sir James Steuart*. London: G. Bell & Sons.

SKINNER, A. S. (1962). Sir James Steuart: economics and politics. *Scottish Journal of Political Economy*, 9, February, pp. 17–37. Rpt. in Blaug (1991), pp. 127–47.

SKINNER, A. S. (1963). Sir James Steuart: international relations. *Economic History Review*, 2nd series, 15, 3, pp. 438–50. Rpt. in Blaug (1991), pp. 148–60.

SKINNER, A. S. (1965). Economics and the problem of method: an eighteenth century view. *Scottish Journal of Political Economy*, 12, November, pp. 267–80. Rpt. in Blaug (1991), pp. 161–74.

SKINNER, A. S. (1981) Sir James Steuart: author of a system. *Scottish Journal of Political Economy*, 28, February, pp. 20–42. Rpt. in Blaug (1991), pp. 274–96.

SKINNER, A. S. (1989). Sir James Steuart: economic theory and policy. In *Philosophy and Science in the Scottish Enlightenment*, P. Jones (ed.). Edinburgh: John Donald.

SKINNER, A. S. (1992). Political economy: Adam Smith and his Scottish predecessors. In *Adam Smith Reviewed*, P. Jones and A. S. Skinner (eds.). Edinburgh: Edinburgh University Press, pp. 217–43.

SKINNER, A. S. (forthcoming). David Hume: economic writings. In the Cambridge edition of Hume's Works.

SMITH, A. (1976a) [1776]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Vol. 2 of the Glasgow Edition of the Works and Correspondence of Adam Smith. Oxford: Clarendon Press.

SMITH, A. (1976b) [1759]. *The Theory of Moral Sentiments*. Vol. 1 of the Glasgow Edition of the Works and Correspondence of Adam Smith, D. D. Raphael and A. L. Macfie (eds.). Oxford: Clarendon Press.

SOMMER, L. (1925). *Die österreichischen Kameralisten in dogmengeschichtlicher Darstellung*, 2nd vol. Vienna: Carl Konegan.

SPIEGEL, H. W. (1987). Hamilton, Alexander. In *The New Palgrave: A Dictionary of Economics*, 2, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

STARK, W. (1968). Steuart, James Denham. *International Encyclopedia for the Social Sciences*.

STETTNER, W. F. (1945). Sir James Steuart on the public debt. *Quarterly Journal of Economics*, 59, May, pp. 451–76. Rpt. in Blaug (1991), pp. 1–26.

STEUART, SIR J. (1966) [1805]. *An Inquiry into the Principles of Political Economy*. Edited [and abridged] by and with a biographical sketch and analytical introduction by A. S. Skinner. Edinburgh: Oliver & Boyd and Chicago: Chicago University Press.

STEUART, SIR J. (1967a) [1805]. *The Works, Political, Metaphysical, and Chronological of Sir James Steuart*, 6 vols. The Adam Smith Library. NY: Augustus M. Kelley.

STEUART, SIR J. (1967b) [1805]. Anecdotes of the life of Sir James Steuart, Baronet. In his *Works*, 6, pp. 361–91.

STEUART, SIR J. (1967c) [1805]. A dissertation on the policy of grain, with a view to a plan for preventing scarcity or exorbitant prices in the common markets of England. In his *Works*, 5, pp. 349–77.

STEUART, SIR J. (1967d) [1761]. A dissertation upon the doctrine and principles of money, applied to the German coin. In his *Works*, 5, pp. 172–265.

TAUTSCHER, A. (1959). 'Kameralismus.' In *Staatslexikon*, 6th ed.

TAYLOR, W. L. (1957). A short life of Sir James Steuart: political economist.' *South African Journal of Economics*, 25, December, pp. 290–302. Rpt. in Blaug (1991), pp. 88–100.

© Scottish Economic Society 1996

TRIBE, K. (1987a). Justi, Johann Heinrich Gottlob von. In *The New Palgrave: A Dictionary of Economics*, 4, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

TRIBE, K. (1987b). List, Friedrich. In *The New Palgrave: A Dictionary of Economics*, 3, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

VAGGI, G. (1987). Physiocrats. In *The New Palgrave: A Dictionary of Economics*, 3, J. Eatwell, M. Mulgate and P. Newman (eds.). London: Macmillan; New York: Stockton; and Tokyo: Maruzen.

VICKERS, D. (1959). *Studies in the Theory of Money 1690–1776*. Philadelphia & NY: Chilton, ch. 12: Sir James Steuart.

VICKERS, D. (1970). Review of *The Works, Political, Metaphysical and Chronological of Sir James Steuart*. *Journal of Economic Literature*, 8, 4, December, pp. 1190–95.

VINER, J. (1937). *Studies in the Theory of International Trade*. New York, London: Harper and Brothers.

WEISWANGE, M. (1923). James Steuarts Einkommenslehre im Vergleich mit den Theorien von Quesnay und Smith. Dissertation, Cologne University.

WEITZEL, J. I. (1832). *Geschichte der Staatswissenschaft*. Stuttgart and Tübingen: Cotta'sche Buchhandlung.

WEST, E. G. (1976). Adam Smith's economics of politics. *History of Political Economy*, 8, 4, pp. 515–39.

WINCH, D. (1978). *Adam Smith's Politics: An Essay in Historiographic Revision*. Cambridge, London, New York, Melbourne: Cambridge University Press.

ZUCKERKANDL, R. (1889). *Zur Theorie des Preiss mit besonderer Berücksichtigung der geschichtlchen Entwicklung der Lehre*. Leipzig: Duncker & Humblot.

Date of receipt of final manuscript: 26 April 1995

© Scottish Economic Society 1996

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

UK EXCESS SHARE RETURNS: FIRM SIZE AND VOLATILITY

*Patricia Fraser**

ABSTRACT

Using the family of GARCH-M(p, q) models and UK data comprising of the market portfolio and a portfolio of smaller company shares over the period January 1970 through June 1994, this paper provides support for the notion that the degree of market capitalisation is an important factor in the analysis of risk-return relationships. The evidence reported supports the view that, although both portfolios appear to be driven by the persistence of volatility shocks, there exist significant differences in risk-return behaviour.

I INTRODUCTION

This paper is concerned with the empirical relationship between the expected excess return on a portfolio, consisting of the return on shares less the return from a relatively risk-free bond (the excess return) and the perceived risk of that portfolio. In particular, emphasis is placed on the analysis of risk-return relationships from the perspective of market capitalisation. Using UK data, comprising the market portfolio and a smaller companies portfolio, the empirical analysis in this paper utilises the family of GARCH-type models to examine whether the expected excess returns of companies with low market capitalisation display similar risk-return characteristics to those of the market as a whole.

A comparison of the market risk-return relationship with that of companies with low market capitalisation is interesting because of evidence which has been reported both in the UK (Dimson and Marsh, 1989; Fraser, 1992), and in the US (Banz, 1981; Roll, 1981, 1983; Reinganun, 1981; Basu, 1983), that small firms have consistently out-performed the market portfolio. In such a scenario, finance theory predicts that the expected excess returns of companies with low market capitalisation will exhibit greater volatility than the excess returns of the market portfolio. The increased volatility reflecting the greater economic uncertainty associated with such shares.

The paper is organised as follows: a brief review of the recent risk-return literature and the econometric methodology used in this study is presented in

* University of Aberdeen.

Section II; Section III describes the data and discusses the empirical results while Section IV concludes.

II Review of Recent Studies and Methodology

While there is a substantial literature on stock market risk-return trade-offs, much of the work undertaken relates to the US rather than to the UK. Such studies, typically measure risk by the volatility of the asset in question and tend to report mixed results for the return-volatility trade-off. Pindyck (1984) for example, reports evidence which suggests that the increase in risk premiums in the US during the 1970s arose from increases in volatility. Poterba and Summers (1988), however, using time series analysis over a similar time period, find evidence which contradicts the original Pindyck findings.

In recent years the analysis of risk-return relationships has been advanced by the development of models which allow expected excess returns to follow a Generalised Autoregressive Conditional Heteroscedasticity (GARCH) process (Bollerslev, 1986). The movement of applied researchers to explicitly model financial time series in terms of second moments is attractive because it captures the propensity of returns to cluster in time and can go some way in explaining the observed non-normality of return distributions.[1] Such models are, by now, well known and have provided good fits for many financial time series (see Bollerslev, 1992 for a useful survey).

For example, French *et al.* (1987), employing a Moving Average and GARCH statistical approach, conclude that expected market excess returns over the longer time period of 1928 through 1984 tend to be significant and positively related to the expected volatility of share returns. The French *et al.* result was supported by Chou (1988) for the period 1965 through 1985 but rejected by Baillie and DeGennaro (1990), over the sample periods 1970 through 1987 and 1928 through 1984.

UK studies on the risk-return relationship are less numerous than those using US data. A recent paper by Poon and Taylor (1992) however, quantifies the risk-return relationship of the market portfolio over the period 1965 through 1989 with the use of sample variances and GARCH models. Using data of different frequency, the authors report that expected returns have a positive though insignificant relationship with expected volatility. Comparison of their results with those from the US however, is difficult, due to the conflicting conclusions of US studies. However, their results suggest that variables other than volatility estimates are used by agents when formulating expected returns.

Overall, it would appear that the specific time period and portfolios studied, as well as the methodology employed, are critical in forming conclusions regarding the risk-return trade-off.

[1] Ross (1989) shows that the variance of changes in speculative asset prices is positively related to the highly non-linear flow of information to the market and may explain the propensity of returns to cluster over time. Further, Lamoureaux and Lastrapes (1990) demonstrate that if 'information arrivals' at the market are serially correlated, this will give rise to a type of persistence which can be detected using a GARCH model.

© Scottish Economic Society 1996

In this paper the conventional GARCH model with conditional normal error distribution is extended to allow for conditional t-distributed errors. Bollerslev (1987) makes the point that such a specification is useful as it allows for any unconditional kurtosis found in the data to be explained by conditional heteroscedasticity and/or a conditional leptokurtic distribution.

The Bollerslev (1987) GARCH$(p, q) \sim t$ model takes the form:

$$y_t = \mu + \varepsilon_t \tag{1}$$

$$\begin{aligned} \varepsilon_t \mid \Omega_{t-1} &\sim f_v(\varepsilon_t \mid \Omega_{t-1}) \\ &= \Gamma(v+1/2)\Gamma(v/2)^{-1}((v-2)h_t^2)^{-1/2} \\ &\quad \times (1 + (\varepsilon_t^2(1/h_t^2))(v-2)^{-1})^{-(v+1)/2}, \quad v>2 \end{aligned} \tag{2}$$

$$h_t^2 = \omega + \sum_{i=1}^{p} \gamma_i h_{t-i}^2 + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}^2 \tag{3}$$

where the conditional mean, y_t, is constant; Ω_{t-1}, denotes information available at time $t-1$; the errors, ε_t, follow a conditional t-distribution; $f_v \sim (\varepsilon_t \mid \Omega_{t-1})$, is the conditional density function for the system (1), (2) and (3); the conditional variance, h_t^2, is a function of last periods conditional variance and also last periods forecast error; degrees of freedom are denoted by, v.[2]

Several specifications of the GARCH$(p, q) \sim t$ model described above[3] were applied to the excess returns of the market portfolio and the smaller company portfolio. Details of the data and the empirical results are discussed in the next section.

III Data Description and Empirical Evidence

Data for this paper were extracted from Datastream and consist of monthly observations on the *Financial Times Actuaries All-Share Index*, the *Hoare-Govett Small Company Index* and the monthly yield from the three month Treasury Bill over the period January 1970 through September 1992.[4]

The small company index comprises approximately 1,200 companies each with a maximum market capitalisation of £100m. The companies make up the lowest 10% of the London International Stock Market (Dimson and Marsh, 1989). For both series, portfolio excess returns were calculated according to the formula:[5]

$$r_{pt} - rf_t \tag{4}$$

where

$$r_{pt} = \ln(p_t/p_{t-1}) \tag{5}$$

[2] In the empirical work below, ω, γ_i and δ_i are constrained to be non-negative.

[3] The model is derived in Bollerslev (1987).

[4] The sample period, 1970–1994, includes values of the variables observed within a system of fixed and floating exchange rates. The analysis therefore covers a period when international factors may have resulted in a particularly volatile equity market.

[5] Returns in this study do not include dividend payments. Given the way dividends are reported, annually and semi-annually, Lo and MacKinlay (1990) suggest they would not add much to the variability of the overall series.

© Scottish Economic Society 1996

and r_{pt} is the return on the portfolio at time t, p_t is the price level of the portfolio the same month and rf_t is the monthly return from the three month Treasury Bill. ln denotes natural logarithms.

Preliminary analysis

Descriptive statistics are reported in Table 1 for both excess return portfolios. The reported sample mean values of the *Financial Times Actuaries All-Share Index* (*FT*) and the *Hoare-Govett Small Companies Index* (*SC*) indicate that the average performance of the smaller company portfolio was greater than that of the market portfolio by 0·7% per month. This would appear to support the UK and US studies discussed above of the existence of a possible 'size effect' in excess returns. Monthly *ex post* standard deviations reveal that the monthly

TABLE 1
Summary statistics for UK stock excess returns[1]

	μ	sd	sk	ku	ARCH(6)		
FT	0·006	0·063	0·289	8·027	24·000		
					(0·000)		
SC	0·013	0·055	−0·474	3·741	14·982		
					(0·020)		
	Autocorrelations $(y-\mu)$ for lags						
	(1)	(2)	(3)	(4)	(5)	(6)	$Q(10)$
FT	0·136	−0·098	0·051	0·055	−0·113	−0·063	16·856
	(0·058)	(0·062)	(0·062)	(0·063)	(0·063)	(0·063)	(0·077)
SC	0·279	0·009	0·061	0·015	−0·062	0·010	30·771
	(0·058)	(0·061)	(0·061)	(0·062)	(0·062)	(0·062)	(0·001)
	Autocorrelations $(y_t-\mu)^2$ for lags						
	(1)	(2)	(3)	(4)	(5)	(6)	$Q^2(10)$
FT	0·138	0·137	0·104	0·063	0·097	0·042	24·210
	(0·058)	(0·059)	(0·060)	(0·061)	(0·061)	(0·061)	(0·007)
SC	0·220	0·034	0·091	0·071	0·133	0·008	27·634
	(0·058)	(0·059)	(0·060)	(0·061)	(0·061)	(0·061)	(0·002)
	Partial autocorrelations $(y-\mu)^2$ for lags						
	(1)	(2)	(3)	(4)	(5)	(6)	
FT	0·138	0·120	0·073	0·027	0·068	0·007	
SC	0·219	−0·015	0·091	0·034	0·117	−0·055	

Notes:
[1] μ, sd, sk and ku denote mean, standard deviation, skewness and kurtosis respectively. sk and ku are from Kendall and Stuart (1958).

ARCH(6) is a test of autoregressive conditional heteroscedasticity for up to six lags. $Q(10)$ and $Q^2(10)$ are the Ljung-Box portmanteau tests for up to tenth order serial correlation in the levels and squares respectively $(y-\mu)$.

Figures in parenthesis under the ARCH and Q statistics are marginal significance levels and under the autocorrelations are standard errors.

FT is the *Financial Times Actuaries All Share Index* and *SC*, *Hoare-Govett Smaller Companies Share Index*.

The sample period is from January 1970 through June 1994.

© Scottish Economic Society 1996

variance of the small company portfolio is considerably less than that of the broadly-based market portfolio with the latter reporting a variance of approximately 0·4% per month compared to approximately 0·3% per month for the SC series. It would appear that the relatively superior performance of the small company portfolio was achieved with less *ex post* total risk (as measured by unconditional standard deviation).

The skewness statistics imply that the *FT* series tends to be positively skewed while the *SC* series tends toward being negatively skewed; from a risk-averse investor's perspective therefore, investing in the broadly-based portfolio would represent a slightly more attractive option since positive skewness implies that the return distribution of the portfolio has a heavier tail of large values hence a higher probability of earning positive returns. In accordance with the GARCH(p, q) model, the kurtosis values imply that both excess return distributions have fatter tails than the corresponding normal distribution. Furthermore, up to sixth-order ARCH effects were detected for both series. The presence of unconditional kurtosis and ARCH in both series suggest one of the GARCH(p, q) family of models may be appropriate.

Significant autocorrelation is detected for both series $(y_t - \mu)$ at a lag of one time period with higher order dependency detected for the smaller company portfolio with $Q(10) = 30{\cdot}771$. The autocorrelations for the squared series $(y_t - \mu)^2$ suggest higher order serial dependence in the squared market portfolio, with $Q^2(10) = 24{\cdot}210$. In general, the autocorrelations coefficients in the squared series appeared to be characterised by more significant values than the autocorrelations in the first moments of the series. Bollerslev (1987) points out that such a characteristic is a feature of GARCH(p, q) models. It should be noted, however, that in the presence of ARCH the detection of autocorrelation is problematic and standard test statistics should be interpreted with care.[6]

There is, therefore, some evidence in the preliminary statistics to suggest a model from the GARCH(p, q) family may be an appropriate model for excess returns. Applying the traditional Box and Jenkins identification techniques to the squared autocorrelations and partial autocorrelations reported in Table 1 would suggest that for the market portfolio the relevant $p + q$ models should not be greater than 4 while, for the smaller company portfolio, a $p + q = 2$ model may be suitable.

Further specification tests are reported in Table 2. Several GARCH$(p, q) \sim t$ model specifications were fitted to the data. Estimates of conditional kurtosis $[3(v-2)(v-4)^{-1}]$ indicate substantial deviations from normality for both series suggesting that a model which allows for t-distributed errors may be appropriate.[7,8] Accordingly, the conditional t-distributed maximum log-

[6] In the presence of ARCH, standard tests of autocorrelation will lead to over-rejections (see Taylor, 1986).

[7] Poon and Taylor (1992) report similar values of v for the market index.

[8] An anonymous referee commented that if $u_t | \Omega_{t-1} \sim N(0, h_t^2)$, then $z_t = u_t/h \sim N(0, 1)$ and z can be tested against the standard normal model. If the kurtosis in this specification is inadequate then another conditional distribution may be appropriate. The statistic, z, was constructed for the p, q specifications detailed in Table 2 and indicated non-trivial departures from normality for both series under consideration.

© Scottish Economic Society 1996

TABLE 2
Maximum log-likelihoods for GARCH(p, q) ~ t[1]

	GARCH(p, q)	Students t-distribution LLF	υ
FT	(1, 1)	570·272	4·796
			(1·086)
	(0, 2)	570·294*	5·053
			(1·289)
	(2, 1)	568·719	5·742
			(1·144)
	(2, 2)	548·047	7·733
			(2·414)
SC	(1, 1)	588·748*	4·133
			(1·239)
	(0, 2)	575·567	4·021
			(1·099)
	(2, 1)	578·776	4·378
			(1·194)
	(2, 2)	574·052	12·687
			(4·294)

Notes:
[1] GARCH(p, q) represent p AR terms and q MA terms in the conditional variance series.
υ is the degrees of freedom for the conditional t-distribution. Figures under the υ statistics are standard errors.
The conservative critical value of the χ^2_1 at the 5% level of significance is 3·84 (see Bollerslev, 1987).
* indicates model with largest value of log-likelihood.

likelihood values are reported for several (p, q) specifications for both series. Standard Likelihood Ratio tests identify a GARCH$(0, 2) \sim t$ model for the market portfolio while for the smaller company portfolio, a GARCH(1, 1) model is identified for estimation.[9]

GARCH-M(p, q) ~ t models

In Tables 3 and 4, GARCH-M(p, q)[10] parameter estimates are reported using two specifications of *ex ante* volatility: conditional variance (α_v) and conditional standard deviation (α_s), respectively. In an attempt to capture any autocorrelation which may be present a moving average term, MA(1), is included in the mean equation. An October 1987 dummy is also included in order to analyse the

[9] Bollerslev (1987) shows that if model 1 is nested within model 2 and the true parameters are in space defined by model 2 then 2(ln 2 – ln 1) approximately follows a χ^2 distribution with $k2 - k1$ degrees of freedom where k is the number of parameters. GARCH(1, 1) results for the market portfolio were estimated but the conditional variance was found to be non-stationary over the period under consideration. The results for the GARCH(1, 1) model are however available from the author upon request.

[10] A GARCH-M(p, q) model is quite plausible. For example, if stock markets are efficient, *ex ante* volatility may be used to help predict excess returns. Accordingly, under the assumption that agents are risk averse, the coefficient on *ex ante* volatility in the mean equation will be significant and positive.

© Scottish Economic Society 1996

© Scottish Economic Society 1996

TABLE 3
Maximum likelihood estimates for the GARCH-M(p, q) ~ t models of excess returns (with MA term and October 1987 dummy)[1]

$$(r_{pt} - rf_t) = \mu + \alpha_v h^2 + \beta D_1 + \varepsilon_t - \theta_{\varepsilon t-1}$$

$$\varepsilon_t | \Omega_{t-1} \sim f_v(\varepsilon_t | \Omega_{t-1})$$

$$h_t^2 = \omega + \sum_{i=1}^{p} \gamma_i h_{t-i}^2 + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}^2$$

	(p, q)	μ	α_v	β	θ	ω	γ_1	δ_1	δ_2	ARCH(6)	LLF	v
FT	(0, 2)	0·008	0·814	−0·040	−0·118	0·049	—	0·368	0·461	13·518	572·152	5·209
		(0·008)	(2·352)	(0·030)	(0·058)	(0·509E-02)	—	(0·187)	(0·100)	(0·036)		(1·696)
SC	(1, 1)	0·014	1·514	−0·136	−0·239	0·101E-02	0·490	0·206	—	13·421	599·418	4·161
		(0·009)	(3·306)	(0·033)	(0·068)	(0·467E-03)	(0·184)	(0·115)		(0·037)		(1·533)

Residual Autocorrelations at lag

	(p, q)	(1)	(2)	(3)	(4)	(5)	(6)	$Q(10)$
FT	(0, 2)	−0·016	−0·132	0·067	0·085	−0·112	−0·054	14·620
								(0·146)
SC	(1, 1)	−0·007	0·018	0·091	0·058	−0·047	0·035	11·019
								(0·356)

Notes:
[1] See notes to Table 1.
β denotes the coefficient on the October 1987 dummy, D_1, and θ the coefficient on the MA term in the mean equation.
Figures in parentheses under the coefficient estimates are standard errors.
For the autocorrelation the estimated standard error is approximately 0·064.

© Scottish Economic Society 1996

TABLE 4
Maximum likelihood estimates for GARCH-M(p, q) ~ t models of excess returns (with MA term and October 1987 dummy)[1]

$$(r_{pt} - rf_t) = \mu + \alpha_z h_t + \beta D_1 + \varepsilon_t - \theta_{tt-1}$$

$$\varepsilon_t | \Omega_{t-1} \sim f_v(\varepsilon_t | \Omega_{t-1})$$

$$h_t^2 = \omega + \sum_{i=1}^{p} \gamma_i h_{t-i}^2 + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}^2$$

	(p, q)	μ	α_z	β	θ	ω	γ_1	δ_1	δ_2	ARCH(6)	LLF	v
FT	(0, 2)	0·003	1·132	0·040	−0·119	0·050	—	0·367	0·463	13·042	572·168	5·224
		(0·021)	(0·369)	(0·030)	(0·059)	(0·509E-02)		(0·187)	(0·101)	(0·042)		(1·708)
SC	(1, 1)	0·049	0·255	−0·138	−0·239	0·993E-03	0·496	0·205	—	14·074	599·477	4·632
		(0·025)	(0·400)	(0·033)	(0·067)	(0·479E-03)	(0·188)	(0·113)		(0·030)		(1·565)
					Residual Autocorrelations at lag							
		(1)	(2)	(3)	(4)	(5)	(6)	$Q(10)$				
FT	(p, q)	0·019	−0·128	0·066	0·082	−0·111	−0·055	14·134				
								(0·167)				
SC	(1, 1)	0·011	0·021	0·092	0·059	−0·046	0·035	11·254				
								(0·338)				

Note:
[1] See notes to Table 3.

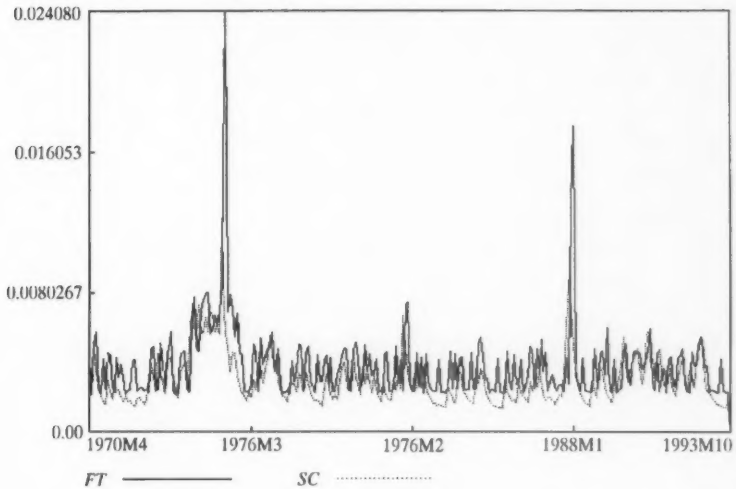


Figure 1. Conditional variance from GARCH-M$(p, q) \sim t$ model.

effect of this outlier on the model. In the standard deviation specification (Table 4) the mean volatility estimate for the market portfolio is positively signed and significant while for the smaller companies portfolio there is no evidence in either of the specifications of a significant conditional variance—mean excess return relationship. The GARCH-M slope parameters are all significant at least at the 10% level of significance and sum to less than unity, which is consistent with a mean reverting volatility process. For both portfolios the MA(1) parameters, θ, are significant but the October dummy, β, is significant at conventional levels only for the smaller company portfolio. Residual ARCH however is significant at approximately 4% and second order serial correlation is present in the residuals of the market portfolio model.

An examination of Figure 1, which plots the conditional variance for both excess return series, shcws that pre 1987, a wedge appears to have been driven between the conditional volatility of the two portfolios. Since 1987 however, small company shares appear to exhibit greater conditional volatility than that of the market which suggests that their relationship with the market portfolio has changed since this event. Figure 1 also implies that the conditional variance for both series increased over periods when there was a substantial drop in share prices (early 1970s and 1987), while in the six-year period prior to 1987, when the stock market was bullish, volatility is less pronounced. Nelson (1991) argues that the sign on excess returns determines future volatility and that the volatility of returns are negatively correlated with the direction of price changes.[11] One

[11] Black (1976) and Christie (1982) also argue that a decline in share prices relative to bond prices has the effect of increasing gearing, therefore, increasing the expected return on the shares in the future and increasing the variance of the stock in the current period.

© Scottish Economic Society 1996

© Scottish Economic Society 1996

TABLE 5

Maximum likelihood estimates for GARCH-M$(p, q) \sim t$ models of excess returns (with MA term, October 1987 dummy and conditional variance dummy)[1]

$$(r_{pt} - rf_t) = \mu + \alpha_v h^2 + \beta D_1 + \varepsilon_t - \theta_{t t-1}$$

$$\varepsilon_t | \Omega_{t-1} \sim f_v(\varepsilon_t | \Omega_{t-1})$$

$$h_t^2 = \omega + \sum_{i=1}^{p} \gamma_i h_{t-i}^2 + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}^2 + \lambda D_{2t-1}$$

	(p, q)	μ	α_v	β	θ	ω	γ_1	δ_1	δ_2	λ	ARCH(6)	LLF	v
FT	(0, 2)	0·004	2·538	−0·042	−0·184	0·046	—	0·292	0·369	0·191E-02	16·174	574·596	4·842
		(0·009)	(2·734)	(0·041)	(0·005)	(0·005)		(0·178)	(0·112)	(0·112E-02)	(0·013)		(1·497)
SC	(1, 1)	0·018	1·588	−0·135	−0·248	0·498E-03	0·604	0·095	—	0·120E-02	13·316	602·604	6·224
		(0·009)	(3·290)	(0·028)	(0·068)	(0·311E-03)	(0·144)	(0·069)		(0·524E-03)	(0·040)		(2·718)

Residual Autocorrelations at lag

	(p, q)	(1)	(2)	(3)	(4)	(5)	(6)	$Q(10)$
FT	(0, 2)	−0·021	−0·117	0·077	0·093	−0·114	−0·070	14·751
								(0·141)
SC	(1, 1)	−0·008	0·011	0·078	0·050	−0·058	0·029	10·017
								(0·439)

Notes:

[1] See notes to Table 3.

λ denotes the coefficient on the conditional variance dummy, $D_{2,t-1}$, where $D_{2,t-1} = 1$ if excess returns $_{t-1} < 0$; 0 otherwise.

© Scottish Economic Society 1996

TABLE 6

Maximum likelihood estimates for GARCH-M(p, q) ~ t models of excess returns (with MA term, and October 1987 dummy and conditional variance dummy)[1]

$$(r_{pt} - rf_t) = \mu + \alpha_s h + \beta D_1 + \varepsilon_t - \theta_{t-1}$$

$$\varepsilon_t | \Omega_{t-1} \sim f_v(\varepsilon_t | \Omega_{t-1})$$

$$h_t^2 = \omega + \sum_{i=1}^{p} \gamma_i h_{t-i}^2 + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}^2 + \lambda D_{2t-1}$$

	(p, q)	μ	α_s	β	θ	ω	γ_1	δ_1	δ_2	λ	ARCH(6)	LLF	v
FT	(0, 2)	−0·013	0·394	−0·041	−0·193	0·046	—	0·283	0·374	0·190E-02	15·560	574·784	4·859
		(0·020)	(0·370)	(0·043)	(0·068)	(0·504E-02)		(0·184)	(0·113)	(0·112E-02)	(0·016)		(1·503)
SC	(1, 1)	0·022	−0·098	−0·135	−0·246	0·498E-03	0·602	0·094	—	0·122E-02	13·146	602·622	6·249
		(0·018)	(0·366)	(0·028)	(0·067)	(0·308E-03)	(0·176)	(0·067)		(0·522E-03)	(0·041)		(2·728)

Residual Autocorrelations

	(p, q)	(1)	(2)	(3)	(4)	(5)	(6)	$Q(10)$
FT	(0, 2)	−0·018	−0·121	0·070	0·090	−0·115	−0·040	13·987
								(0·174)
SC	(1, 1)	−0·009	0·009	0·077	0·049	−0·059	0·028	9·921
								(0·447)

Note:
[1] See notes to Table 5.

simple way to test this hypothesis is to include a lagged intercept dummy in the conditional variance equation which will take a value of unity when returns are less than zero and a value of zero otherwise. Essentially, if the coefficient on the lagged intercept dummy (λ) is significantly positive then this indicates a marked upward shift in the conditional variance when returns are negative. The result of this test using both specifications of volatility is reported in Tables 5 and 6.[12]

Considering the market portfolio first, λ, is insignificantly different from zero but log-likelihood values have improved (compared to those reported in Tables 3 and 4). The mean conditional volatility parameter, α, remains positive but is now insignificant in both market portfolio specifications. There remains evidence of residual ARCH at the 5% level but residual autocorrelation is rejected in the variance specification of this model (Table 5) although in the standard deviation specification (Table 6), some residual dependency is still evident at lag 2. In both specifications, the GARCH slope coefficients sum to less than unity and the coefficient, δ_2, is significant.

For the smaller company portfolio, λ, is significant: in contrast with the market portfolio, it would appear that 'when stocks go up, volatilities seem to go down: and when stocks go down, volatilities seem to go up', Black (1976). The log-likelihood value has also increased and the coefficients on the October 1987 remain significant. While the mean conditional volatility parameter is insignificant it has a negative value in both specifications. The conditional variance appears to be stationary and the autoregressive conditional variance parameter, γ_1, is significant for both specifications of the model. As the magnitude and significance of the moving average conditional variance parameter, γ_2, is now considerably less than that reported in Tables 3 and 4, the implication is that market conditions may be a source of the observed persistence in the excess returns of the small company portfolio. There is no sign of dependency in the residuals of the models and ARCH is rejected at the 4% level of significance.

IV Conclusion

Using UK data for the period January 1970 through June 1994, this paper examines the risk-return relationships of a small company portfolio and a broadly-based market portfolio. Using the family of GARCH-M$(p, q)\sim t$ models to derive measures of conditional volatility, the empirical work tests the hypothesis that the expected excess returns of companies with low market capitalisation display similar risk-return characteristics to those of the market as a whole.

The empirical findings suggest that, over the period under consideration, the excess returns for both portfolios are driven by the persistence of volatility shocks. The evolution of the volatility processes would, however, appear to be different. The stock market crash of October 1987 would also appear to have

[12] This test was suggested by an anonymous referee.

© Scottish Economic Society 1996

had a greater effect on the smaller company portfolio than that of the market portfolio. Interestingly, there is some, albeit tenuous, evidence to suggest that, post 1987, smaller company excess returns have exhibited relatively greater *ex ante* volatility than the market as a whole. The suggestion is that smaller company shares may have become more actively traded in the post October 1987 period. For the smaller company portfolio in particular, volatility shocks appear to have no effect on the mean value of the portfolio. It is interesting to note, however, that while the relationship is positive for the market portfolio, it is negative for the small company portfolio, implying that small company investors may have greater risk loving characteristics than the average market investor. Further, the conditional variance of the portfolios respond in a different manner to bullish and bearish conditions in the market for their shares. The expected volatility of smaller company shares appearing to be more sensitive to movements in their stock prices than that of the market as a whole.

Overall, the evidence reported in this paper would suggest, as we would expect, that smaller company shares and the market as a whole have some common characteristics. There exists, however, significant differences in risk-return behaviour. The source of differences in risk-return behaviour may be because information on smaller companies, has in the past, tended not to be available to all traders simultaneously and also less likely to be acted upon immediately. The collection and analysis of small company data of higher frequencies and over a wider range of sample periods may be informative in this respect. Similarly, modelling the GARCH process to include the flow of information to the market may also prove to be worthwhile. These topics provide ample scope for further research.

ACKNOWLEDGEMENTS

The author wishes to thank Professor R. MacDonald, Dr D. Power and Dr N. Groenewold for their comments on an earlier draft of this paper. She also acknowledges the constructive comments of two anonymous referees.

REFERENCES

BAILLIE, R. T. and DEGENNARO, R. P. (1990). Stock returns and volatility. *Journal of Financial and Quantitative Analysis*, 25, pp. 302–215.

BANZ, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, pp. 3–18.

BASU, S. (1983). Investment performance of common stocks in relation to their price-earnings ratio: a test of the efficient markets hypothesis. *Journal of Finance*, pp. 663–682.

BLACK, F. (1976). Studies of stock price volatility changes. *Proceeds of the American Statistical Association, Business and Economics Statistics Section*, pp. 177–181.

BOLLERSLEV, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, pp. 307–328.

BOLLERSLEV, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 59, pp. 542–547.

© Scottish Economic Society 1996

BOLLERSLEV, T., CHOU, R. Y. and KRONER, K. F. (1992). Arch modeling in finance. *Journal of Econometrics*, 52, pp. 5–59.

CHOU, R. Y. (1988). Volatility persistence and stock evaluations: some empirical evidence using GARCH. *Journal of Applied Econometrics*, 3, pp. 279–294.

CHRISTIE, A. A. (1982). The stochastic behaviour of common stock variances: value, leverage and interest rate effects. *Journal of Financial Economics*, 10, pp. 407–432.

DIMSON, E. and MARSH, P. (1989). The smaller companies puzzle. *The Investment Analyst*, 91, pp.16–24.

ENGLE, R. F. and BOLLERSLEV, T. (1986). Modelling the persistence of conditional variances. *Econometric Review*, 5, 1, pp. 1–50.

FRASER, P. (1995). Returns and firm size: a note on the UK experience 1970–1991. Forthcoming *Applied Economics Letters*.

FRENCH, K. R., SCHWERT, G. W. and STAMBAUGH, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 9, pp. 3–29.

LAMOUREAUX, C. G. and LASTRAPES, W. D. (1990). Heteroskedasticity in stock return data: volume versus Garch effects. *The Journal of Finance*, 45, pp. 221–229.

LO, A. and MACKINLAY, C. (1990). When are contrarian profits due to stock market overreaction. *Review of Financial Studies*, 1, pp. 175–205.

NELSON, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59, pp. 347–370.

PINDYCK, R. S. (1984). Risk, inflation, and the stock market. *American Economic Review*, 74, pp. 335–351.

POON, S. H. and TAYLOR, S. (1992). Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance*, 16, pp. 37–59.

POTERBA, J. M. and SUMMERS, L. H. (1986). The persistence of volatility and stock market fluctuations. *American Economic Review*, 76, pp. 1142–1151.

REINGANUM, M. R. (1981). Misspecification of capital asset pricing. Empirical anomalies based on earnings' yields and market values. *Journal of Financial Economics*, 9, pp. 19–46.

ROLL, R. (1981). Possible explanation of the small firm effect. *Journal of Finance*, pp. 879–888.

ROLL, R. (1983). On computing mean returns and the small firm premium. *Journal of Financial Economics*, pp. 371–386.

ROSS, S. A. (1989). Information and volatility: the no-arbitrage martingale approach to timing and resolution irrelevancy. *Journal of Finance*, 44, pp. 1–17.

TAYLOR, S. (1986). *Modelling Financial Time Series*, John Wiley and Sons Ltd.

Date of receipt of final manuscript: 13 February 1995

© Scottish Economic Society 1996

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

SUNK COSTS AND PROFIT TAXATION: A SOURCE OF TAX NON-NEUTRALITY

*Paolo M. Panteghini**

ABSTRACT

This paper uses a two-period Von Stackelberg model to study the effects of profit taxation on the behaviour of a monopolist, when the entry of a potential competitor is threatened. A barrier to entry, consisting of a sunk cost, is assumed. If the potential competitor decides to enter, thus making a loss, deduction is allowed in the following period. This model shows that these tax deductions can make profit taxation be distortive.

I INTRODUCTION

Many works have dealt with distortions caused by corporation taxes, in particular analysing commodity taxation. Despite the importance of profit taxation, only a few have dealt with its effects in an imperfect market. In the history of fiscal doctrine, many economists have agreed that in a monopolistic market a profit tax cannot be shifted.[1] This kind of tax can in fact be distortive.

As Musgrave and Musgrave (1984) argue, 'profit maximization of this type may not be a realistic interpretation of how firms behave' (p. 280). They suggest some possible causes of distortion of monopolistic behaviour, such as price restraints, wage demands under collective bargaining and sales maximisation rather than output. They also consider the interaction between oligopolistic firms operating in the same market as a further cause of tax distortion. Thus, they refer to an emerging stream of literature, which, by incorporating game theory principles, deals with the links between corporate taxation and market structure (with a whole range of intermediate cases between perfect competition and pure monopoly).

In their fundamental work, Atkinson and Stiglitz (1980) show that profit taxation affects the behaviour of monopolistically competitive firms. However, only a few economists, including Stern (1987) and Davidson and Martin (1991), have followed this example.

[1] See Musgrave (1985), who quotes eminent economists, such as Marshall (1890) and Wicksell (1896). See also Musgrave (1959), p. 278, footnote 1, for an analytical demonstration of this neutrality.

* Università degli Studi di Brescia, Italy.

Stern (1987) studies a particular sales tax rate, which 'is like a tax on profits before fixed cost, with a marginal cost decreed by the government' (p. 142). By assuming 'a spectrum of possible market structures and outcomes, including perfect competition, monopolistic competition, Cournot-oligopoly and monopoly ...' (p. 134), he shows that this particular kind of tax affects prices, profits and welfare, in the cases of both free and blocked entry.

Under the hypothesis that sunk costs create a barrier to entry, Davidson and Martin (1991) find that 'with full loss offset for sunk costs, a pure profit tax cannot affect the entry decision. However, if losses are only partially deductible against future income, the tax will have an impact on the decision to enter' (p. 183).

The works by Atkinson and Stiglitz (1980) and Stern (1987) are based on a static framework, while Davidson and Martin (1991) use an atemporal model, comparing only steady-state equilibria. Despite their important contributions, therefore, the above works do not exhaust all the possible effects of profit taxation, as the time variable is not considered.

By using a two-period leader-follower model, this work introduces a simple transition path allowing the study of the short-term effects of profit taxation. A standard Von Stackelberg model is used, where firm 1 (the incumbent) chooses the production level and firm 2 (the potential entrant) observes this quantity and decides whether to enter the market. Apart from a sunk cost, paid when firm 2 decides to enter, no other costs exist. By assumption there is no capital accumulation.

This framework is helpful for studying the effects of tax deductions. In a static model, profit must be non-negative, otherwise a firm would not produce. Conversely, in an intertemporal context, a firm's short-term results may be negative, if they are offset in the long run. Thus, firm 2 can decide to enter, even making a loss, if it is offset by the second period profits.

Fiscal systems usually allow partial or total deduction of losses against future profits. Therefore, if firm 2 makes a loss during the first period (in which entry occurs), this loss can be deducted against the second period profits. However, if there is a positive real interest rate, the lower the present value of this fiscal benefit, the higher the interest rate. As will be shown, the discount of this fiscal benefit introduces a distortion: when the interest rate is positive, a change in the profit tax rate can modify the behaviour of the entrant and consequently that of the incumbent.

The remainder of the paper is structured as follows. Section II develops the model. Section III discusses the effects of profit taxation on output, while Section IV deals with the effects of this tax on market price and economic welfare. Finally, Section V summarises the results of the work.

II THE MODEL: A LEADER-FOLLOWER FRAMEWORK WITH PROFIT TAXATION

This section develops the two-period Von Stackelberg model. Let the demand

© Scottish Economic Society 1996

for the good produced be linear:[2]

$$p_t = 1 - q_t^1 - q_t^2,$$

where p_t is the market price and q_t^1 and q_t^2 are, respectively, the output of the leader (incumbent) and the follower (entrant), at time $t \in [1, 2]$. The entrant maximises the following profit function:

$$P^2 = (((1 - q_1^1 - q_1^2)q_1^2 - f) - \tau \max(0, (1 - q_1^1 - q_1^2)q_1^2 - f)) + \frac{1}{1+r}$$
$$\times((1 - q_2^1 - q_2^2)q_2^2 - \tau((1 - q_2^1 - q_2^2)q_2^2$$
$$+\min(0, (1 - q_1^1 - q_1^2)q_1^2 - f))), \qquad (1)$$

where τ is the profit tax rate, r is the positive real interest rate and f is a given sunk cost. Function (1) shows that if the follower decides to enter and makes a loss, then the loss will be deducted in the second period. If the entrant makes a profit, during the first period, obviously no deductions will be made.[3]

By assumption, if $P^2 \leq 0$, the follower does not enter. Note that P^2 is non-positive only if $((1 - q_1^1 - q_1^2)q_1^2 - f)$ is negative. This condition is necessary but not sufficient. In fact, even if $((1 - q_1^1 - q_1^2)q_1^2 - f)$ were negative, P^2 could be positive because the entrant could obtain a fiscal benefit, by deducting in the second period the losses made during the first one. Under this condition, the profit function of the entrant becomes:

$$P^2 = \frac{1+r-\tau}{1+r}((1 - q_1^1 - q_1^2)q_1^2 - f) + \frac{1-\tau}{1+r}(1 - q_2^1 - q_2^2)q_2^2. \qquad (1')$$

Maximising function (1′) yields the following first order conditions:

$$\frac{\partial P^2}{\partial q_1^2} = \left(1 - \frac{\tau}{1+r}\right)(1 - q_1^1 - 2q_1^2) = 0,$$

$$\frac{\partial P^2}{\partial q_2^2} = \left(\frac{1-\tau}{1+r}\right)(1 - q_2^1 - 2q_2^2) = 0,$$

which yield $q_1^{2*} = (1 - q_1^1)/2$ and $q_2^{2*} = (1 - q_2^1)/2$. Substituting q_1^{2*} and q_2^{2*} into function (1′) and rearranging the inequality $P^2 \leq 0$ gives the following constraint:

$$\mathcal{H}(r, \tau, f, q_1^1, q_2^1) = \frac{1}{f}\left(\frac{1-q_1^1}{2}\right)^2 + \frac{1-\tau}{1+r-\tau}\frac{1}{f}\left(\frac{1-q_2^1}{2}\right)^2 - 1 \leq 0. \qquad (2)$$

[2]Obviously, the linear demand function can be interpreted as the linearisation of a more general function around an equilibrium point.

[3]Note that the deduction of losses is allowed for both the agents. However, the incumbent cannot exploit this fiscal benefit since he faces no losses.

© Scottish Economic Society 1996

Inequality (2) is the condition under which P^2 is non-positive. Since the aim of the incumbent is to prevent the entry of firm 2, he maximises his profit function under constraint (2):

$$\max_{q_1^1, q_2^1} P^1 = (1-\tau)\left((1-q_1^1-q_1^2)q_1^1 + \frac{1}{1+r}(1-q_2^1-q_2^2)q_2^1\right) \tag{3}$$

$$\text{s.t.} \quad \frac{1}{f}\left(\frac{1-q_1^1}{2}\right)^2 + \frac{1-\tau}{1+r+\tau}\frac{1}{f}\left(\frac{1-q_2^1}{2}\right)^2 - 1 \leqslant 0,$$

where $q_1^2 = q_2^2 = 0$, because of the existence of the constraint.[4] The Lagrangian becomes:

$$\mathcal{L} = (1-\tau)\left((1-q_1^1)q_1^1 + \frac{1}{1+r}(1-q_2^1)q_2^1\right) +$$

$$+\lambda\left(1 - \frac{1}{f}\left(\frac{1-q_1^1}{2}\right)^2 - \frac{1-\tau}{1+r-\tau}\frac{1}{f}\left(\frac{1-q_2^1}{2}\right)^2\right). \tag{4}$$

In this case the first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial q_1^1} = (1-\tau)(1-2q_1^1) + \frac{\lambda}{2f}(1-q_1^1) = 0, \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial q_2^1} = \frac{1-\tau}{1+r}(1-2q_2^1) + \frac{\lambda}{2f}\frac{1-\tau}{1+r-\tau}(1-q_2^1) = 0, \tag{6}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \frac{1}{f}\left(\frac{1-q_1^1}{2}\right)^2 - \frac{1-\tau}{1+r-\tau}\frac{1}{f}\left(\frac{1-q_2^1}{2}\right)^2 \geqslant 0, \quad \lambda \geqslant 0 \tag{7}$$

and

$$\lambda\left(1 - \frac{1}{f}\left(\frac{1-q_1^1}{2}\right)^2 - \frac{1-\tau}{1+r-\tau}\frac{1}{f}\left(\frac{1-q_2^1}{2}\right)^2\right) = 0.$$

Consider the case in which the constraint is slack. When $\mathcal{H}(r, \tau, f, q_1^1, q_2^1) \leqslant 0$

[4] The other constraint

$$(1-q_1^1-q_1^{2^*})\cdot q_1^{2^*} = \left(1-q_1^1-\frac{1-q_1^1}{2}\right)\cdot\frac{1-q_1^1}{2} \leqslant f$$

can be omitted, since it is never binding.

© Scottish Economic Society 1996

and $\lambda = 0$, the above first order conditions become:

$$\frac{\partial P^1}{\partial q_1^1} = (1-\tau)(1-2q_1^1) = 0, \tag{5'}$$

$$\frac{\partial P^1}{\partial q_2^1} = \frac{1-\tau}{1+r}(1-2q_2^1) = 0, \tag{6'}$$

$$\lambda = 0, \tag{7'}$$

which yield the monopolistic solution $q_1^1 = q_2^1 = 1/2$. Under these conditions, taxation is neutral, since the optimal levels of q_1^1 and q_2^1 do not depend on τ. In this case, profit is equal to

$$\frac{1}{4}\,\frac{(1-\tau)(2+r)}{1+r}.$$

The slackness of the constraint can be observed from another point of view. Solving the constraint for f, given the monopolistic choice $q_1^1 = q_2^1 = 1/2$, yields

$$f \geq f^* \equiv \frac{1}{16}\,\frac{2+r-2\tau}{1+r-\tau}.$$

In other words, the slackness of the constraint means that f is greater than the upper limit f^*, above which entry is blocked.

Consider the second case, when $\mathcal{H}(r, \tau, f, q_1^1, q_2^1) = 0$ and $\lambda > 0$, i.e. $f < f^*$. Substituting condition (7) into conditions (5) and (6) and solving for q_1^1 and q_2^1 respectively yields

$$q_1^1 = \frac{(1-\tau) + \dfrac{\lambda}{2f}}{2(1-\tau) + \dfrac{\lambda}{2f}} > \frac{1}{2}$$

and

$$q_2^1 = \frac{1 + \dfrac{1+r}{1+r-\tau} \cdot \dfrac{\lambda}{2f}}{2 + \dfrac{1+r}{1+r-\tau} \cdot \dfrac{\lambda}{2f}} > \frac{1}{2}.$$

As can be seen, both q_1^1 and q_2^1 are greater than the monopolistic level $1/2$. This means that when the constraint is binding and the incumbent does not accommodate, he has to produce more than the monopolistic level, thus reducing prices in both periods.

When f is low enough, however, it might be convenient for the monopolist to accommodate. The lower limit of f, under which accommodation takes place,

© Scottish Economic Society 1996

can be computed as follows. Substituting condition (5) into condition (6), to eliminate λ, and solving for q_2^1 yields:

$$q_2^1 = \mathcal{G}(r, \tau, q_1^1) = 1 - \frac{1 - q_1^1}{2(1 - q_1^1) + \dfrac{(1+r)(1-\tau)}{(1+r-\tau)}(2q_1^1 - 1)}. \tag{8}$$

Now substituting solutions $q_1^{2*} = (1 - q_1^1)/2$, $q_2^{2*} = (1 - q_2^1)/2$ and equation (8) into the general profit function (3) yields the monopolistic profit, in the entry constrained case

$$P^1 = (1 - \tau) \cdot \left(\frac{1 - q_1^1}{2} q_1^1 + \frac{1 - \mathcal{G}(r, \tau, q_1^1)}{2(1+r)} \mathcal{G}(r, \tau, q_1^1) \right), \tag{9}$$

which is increasing in f. This follows because derivative $(\mathrm{d}/\mathrm{d}f)P^1(q_1^1, q_2^1, 0, 0)$ is always positive, as shown in Appendix A.

The monopolistic profit (9) has to be compared with the profit obtained if accommodation occurs

$$P^1\left(q_1^1 = \frac{1}{2}, q_2^1 = \frac{1}{2}, q_1^2 = \frac{1}{4}, q_2^2 = \frac{1}{4}\right) = \frac{1}{8} \frac{(1-\tau)(2+r)}{1+r}.$$

If $P^1(q_1^1, q_2^1, 0, 0) > (1/8)(1 - \tau)(2 + r)/(1 + r)$, then accommodation does not take place. Otherwise, if the opposite is true, accommodation is convenient. Therefore, the value of f which makes the choices between accommodation and non-accommodation indifferent may be defined as follows

$$f^{**} = f \mid P^1(q_1^1, q_2^1, 0, 0) - \frac{1}{8} \frac{(1-\tau)(2+r)}{1+r} = 0.$$

As shown in Figure 1, for $f \leqslant f^*$, two sub-cases follow: $f \in [f^{**}, f^*]$ and $f \in [0, f^{**})$. The equilibrium is one of non-entry if $f \in [f^{**}, f^*]$.

As can be seen, for $f \in [0, f^{**})$, the incumbent's profit is equal to

$$\frac{1}{8} \frac{(1-\tau)(2+r)}{1+r} > P^1(q_1^1, q_2^1, 0, 0).$$

Therefore, a duopoly takes place, where the leader produces $q_1^1 = q_2^1 = 1/2$ and the follower $q_1^2 = q_2^2 = 1/4$. As a consequence, production is not influenced by profit taxation.

When $f \in [f^{**}, f^*]$, P^1 increases as f increases, until the sunk cost is equal to f^*. In this sub-case production depends on the profit tax rate, thus causing taxation to be non-neutral. Finally, for $f \in (f^*, +\infty)$, P^1 is equal to the monopolistic profit and production is not influenced by profit taxation. To summarise, there are two regions ($f \in [0, f^{**})$ and $f \in (f^*, +\infty)$) in which taxation is neutral and one region ($f \in [f^{**}, f^*]$) in which it influences the quantity produced.

© Scottish Economic Society 1996

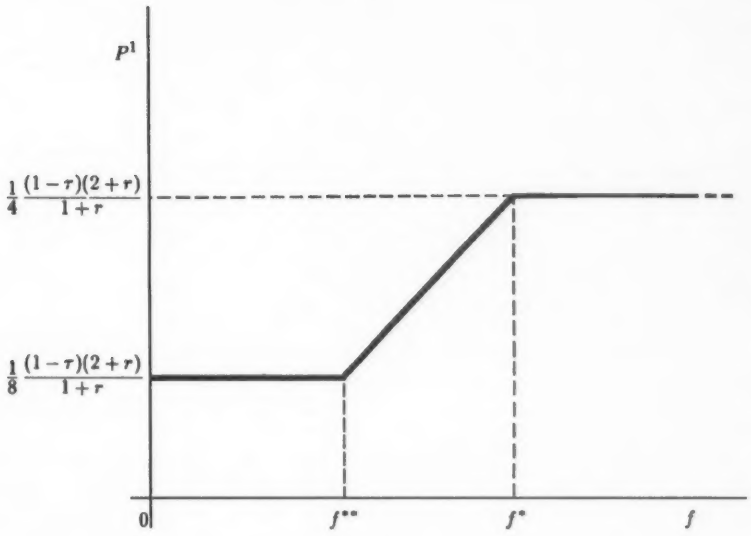


Figure 1.

III THE EFFECTS OF A PROFIT TAX RATE CHANGE ON PRODUCTION LEVEL

In this section the effects of a small profit tax rate change will be studied in the non-neutral region $f \in [f^{**}, f^{*}]$. Under the assumption that r does not change, differentiating constraint (2) yields:

$$\mathcal{H}_{q_2^1}\, dq_2^1 + \mathcal{H}_{q_1^1}\, dq_1^1 + \mathcal{H}_{\tau}\, d\tau = 0, \tag{10}$$

where

$$\frac{\partial \mathcal{H}}{\partial q_1^1} = \mathcal{H}_{q_1^1} = -\frac{1}{2f}(1 - q_1^1) < 0,$$

$$\frac{\partial \mathcal{H}}{\partial q_2^1} = \mathcal{H}_{q_2^1} = -\frac{1-\tau}{1+r-\tau}\,\frac{1}{2f}(1 - q_2^1) < 0,$$

$$\frac{\partial \mathcal{H}}{\partial \tau} = \mathcal{H}_{\tau} = -\frac{r}{(1+r-\tau)^2}\,\frac{1}{4f}(1 - q_2^1) < 0.$$

By using the derivative of equation (8) (see (A3) in Appendix A) and (10), three propositions follow.

Proposition 1. *If $f \in [f^{**}, f^{*}]$, the effect of a profit tax rate change on the quantity produced during the first period depends on the starting outcome:*

$$\operatorname{sign}\left(\frac{dq_1^1}{d\tau}\right) = \operatorname{sign}(q_1^1 - q_1^{1*}).$$

© Scottish Economic Society 1996

Proof. Substituting (A3) into equation (10) to eliminate dq_2^1 and solving the above system for dq_1^1 yields:

$$\frac{dq_1^1}{d\tau} = \frac{\mathcal{H}_{q_2^1}\mathcal{G}_\tau + \mathcal{H}_\tau}{\mathcal{H}_{q_2^1}\mathcal{G}_{q_1^1} + \mathcal{H}_{q_1^1}}, \tag{11}$$

which is positive for

$$q_1^1 \geq q_1^{1*} = \frac{1 + \frac{1}{2}\frac{(1-\tau)(1+r)}{1+r-\tau}}{1 + \frac{(1-\tau)(1+r)}{1+r-\tau}} > \frac{1}{2}$$

and negative for lower values. Therefore proposition 1 is proven.

Q.E.D.

An explanation for this ambiguity can be found in the behaviour of the potential entrant. Profit taxation produces two opposing effects on the entrant's resources. The first is due to the reduction of resources, caused by the profit tax itself. The second is due to benefits arising from the deduction of sunk costs. Thus, a higher profit tax rate increases the reduction of these resources, while simultaneously increasing the benefits of sunk cost deductions. For this reason, if the reduction effect (RE) is larger than the deduction effect (DE), an increase in τ makes the entry of the potential competitor more difficult. In this case, the incumbent can reduce the production level, thus approaching the monopolistic one. If RE is lower than DE, the tax rate increase is an incentive for the follower to enter. The incumbent must increase output, to prevent entry, thereby offsetting this incentive.

Furthermore, the incumbent faces two opposing tax effects, as can be shown in the numerator of derivative (11). There is a direct (negative) effect on the constraint $\mathcal{H}_\tau$, and an indirect (positive) effect on the constraint through $q_2^1, \mathcal{H}_{q_2^1}\mathcal{G}_\tau$. Therefore, when $q_1^1 < q_1^{1*}$, the direct effect $|\mathcal{H}_\tau|$ is larger than the indirect one. Thus, an increase (decrease) in taxation causes q_1^1 to fall (rise). When $q_1^1 > q_1^{1*}$ (and the direct effect is less than the indirect one), the converse is true. The results can be summarised as follows:

$$\operatorname{sign}\left(\frac{dq_1^1}{d\tau}\right) = \operatorname{sign}(DE - RE) = \operatorname{sign}(q_1^1 - q_1^{1*}) = (\mathcal{H}_{q_2^1}\mathcal{G}_\tau - |\mathcal{H}_\tau|).$$

Consequently, the reduction effect of the entrant is related to the direct effect of τ on the constraint faced by the incumbent, $|\mathcal{H}_\tau|$, while the deduction effect is related to the indirect effect, $\mathcal{H}_{q_2^1}\mathcal{G}_\tau$.

Proposition 2. *If* $f \in [f^{**}, f^*]$, *derivative* $dq_2^1/d\tau$ *is always negative.*

Proof. Substituting derivative (11) into derivative (A3) and rearranging it

© Scottish Economic Society 1996

yields:

$$\frac{dq_2^1}{d\tau} = \frac{\mathcal{H}_{q_1^1}\mathcal{G}_\tau - \mathcal{G}_{q_1^1}\mathcal{H}_\tau}{\mathcal{H}_{q_2^1}\mathcal{G}_{q_1^1} + \mathcal{H}_{q_1^1}} < 0 \tag{12}$$

which is unambiguously negative.

Q.E.D.

In other words, if the potential competitor decided to enter in the second period, no deductions could be made, thus making entry non-convenient. Thus, quantity q_2^1 is the output produced by the incumbent, in the second period, when the potential competitor was previously prevented from entering. Hence, an increase in τ will unambiguously reduce q_2^1.

Note that a necessary condition for the existence of distortive effects is the positive value of the interest rate r, which reduces the tax benefit due to the deduction of sunk costs, $\tau f/(1+r)$ (which is less than τf). As can easily be seen, setting $r=0$ yields $\mathcal{G}_\tau = \mathcal{H}_\tau = 0$, thus causing both $dq_1^1/d\tau$ (since the equality RE = DE always holds) and $dq_2^1/d\tau$ to be null. Consequently, taxation is also neutral in the region $[f^{**}, f^*]$.

Proposition 3. *Let Q be total production of the monopolist: $Q = \sum_{i=1}^{2} q_i^1$. If $f \in [f^{**}, f^*]$, then derivative $dQ/d\tau$ is always negative.*

Proof. Differentiating Q and using differential equations (11) and (12) yields:

$$dQ = \frac{(\mathcal{H}_{q_1^1} - \mathcal{H}_{q_2^1})\mathcal{G}_\tau - \mathcal{H}_\tau(1 + \mathcal{G}_{q_1^1})}{\mathcal{H}_{q_2^1}\mathcal{G}_{q_1^1} + \mathcal{H}_{q_1^1}}\, d\tau. \tag{13}$$

Since $\mathcal{H}_{q_1^1} - \mathcal{H}_{q_2^1}$ is always negative (for $q_1^1 \geq 0$), as can be seen in Appendix B, the sign of $dQ/d\tau$ is always negative.

Q.E.D.

Derivative $dq_1^1/d\tau$ can be considered as a short-term effect and $dQ/d\tau$ as a long-term one. Therefore, when $q_1^1 \geq q_1^{1*}$, we have

$$\text{sign}(dq_1^1) = -\text{sign}(dQ).$$

Similar results have been found in a non-strategic context by Auerbach (1989). By assuming that fiscal policy is previously announced, he explains why current investment moves 'in a different direction than that suggested by the long-run effects' (p. 969). Since capital is the only input employed, Auerbach can also explain the opposing short-term direction of output.[5] Conversely, this

[5] Auerbach (1989) explains this short-term effect by introducing a particular investment unit cost function, depending on both investment and capital. See also Panteghini (1995), who finds the same results by eliminating Auerbach's particular technological hypothesis and assuming that the firm under study has monopolistic power.

© Scottish Economic Society 1996

model shows that these opposing signs can also be explained by a simple strategic model.

IV THE EFFECT OF TAXATION ON MARKET PRICE AND ECONOMIC WELFARE

The effects of profit taxation on market price and welfare, when $f \in [f^{**}, f^{*}]$, will be considered. Given propositions 1 and 2, the following corollary holds.

Corollary. *Suppose that $f \in [f^{**}, f]$. During the first period, a profit tax rate change causes the following effect on the market price:*

$$\operatorname{sign}\left(\frac{\mathrm{d}p_1}{\mathrm{d}\tau}\right) = -\operatorname{sign}(q_1^1 - q_1^{1*}).$$

Conversely, the effect of taxation on the second period market price, p_2, is always as follows

$$\operatorname{sign}\left(\frac{\mathrm{d}p_2}{\mathrm{d}\tau}\right) = -\operatorname{sign}\left(\frac{\mathrm{d}q_2^1}{\mathrm{d}\tau}\right).$$

During the first period, the effect of the fiscal change on the market price is ambiguous, depending on the value of q_1^1. When it is sufficiently large, taxation is not shifted into price, because derivative $\mathrm{d}p_1/\mathrm{d}\tau$ is negative. When it is small enough, the converse is true. As regards the second period, derivative $\mathrm{d}p_2/\mathrm{d}\tau$ is always positive, thus causing taxation to be shifted into the market price.

Let us now define welfare as consumers' surplus plus profits. The following proposition holds.

Proposition 4. *Define welfare as follows:*

$$W = \sum_{t=1}^{2} \left(\frac{1}{1+r}\right)^{t-1} w_t$$

where

$$w_t \equiv \int_0^{q_t^1} p_t(u)\,\mathrm{d}u = \int_0^{q_t^1} (1 - q_t^1(u))\,\mathrm{d}u.$$

If $f \in [f^{**}, f^{*}]$, a change in τ yields:

$$\text{(i)}\ \operatorname{sign}\left(\frac{\mathrm{d}w_1}{\mathrm{d}\tau}\right) = \operatorname{sign}(q_1^1 - q_1^{1*})$$

$$\text{(ii)}\ \frac{\mathrm{d}w_2}{\mathrm{d}\tau} < 0;$$

$$\text{(iii)}\ \frac{\mathrm{d}W}{\mathrm{d}\tau} < 0.$$

© Scottish Economic Society 1996

Proof. Differentiating W with respect to τ yields:

$$\frac{dW}{d\tau} = \sum_{t=1}^{2} \left(\frac{1}{1+r}\right)^{t-1} \frac{dw_t}{d\tau}, \tag{14}$$

where $dw_t/d\tau = (1 - q_t^1)\, dq_t^1/d\tau$. Now using propositions 1 and 2 gives:

$$\text{sign}\left(\frac{dw_1}{d\tau}\right) = \text{sign}\left(\frac{dq_1^1}{d\tau}\right) = \text{sign}(q_1^1 - q_1^{1*})$$

and

$$\frac{dw_2}{d\tau} = (1 - q_2^1)\, \frac{dq_2^1}{d\tau} < 0.$$

Consider now point (iii) of the proposition. Substituting equation (8) and derivatives (11) and (12) into derivative (14) yields

$$\frac{dW}{d\tau} = -\frac{1 - q_1^1}{\mathcal{H}_{q_2^1}\mathcal{G}_{q_1^1} + \mathcal{H}_{q_1^1}} \times \left(\left(1 + \frac{\mathcal{G}_{q_1^1}}{(1+r)\mathcal{F}(r, \tau, q_1^1)}\right)\mathcal{H}_\tau + \left(\mathcal{H}_{q_2^1} - \frac{\mathcal{H}_{q_1^1}}{(1+r)\mathcal{F}(r, \tau, q_1^1)}\right)\mathcal{G}_\tau\right)$$

where

$$\mathcal{F}(r, \tau, q_1^1) = 2(1 - q_1^1) + \frac{(1+r)(1-\tau)}{(1+r-\tau)}\,(2q_1^1 - 1),$$

which is always negative, since it is easy to show that

$$\mathcal{H}_{q_2^1} - \frac{\mathcal{H}_{q_1^1}}{(1+r)\mathcal{F}(r, \tau, q_1^1)}$$

is always positive.

Q.E.D.

Therefore, the effect of the first period is ambiguous. Quite surprisingly, when $q_1^1 > q_1^{1*}$, an increase in profit tax can cause an increase in economic welfare.[6] However, this is only a short-term effect. In the long run, fiscal imposition always reduces economic welfare.

[6] In terms of a commodity tax rate change, a similar result has been found by Besley (1989). According to his results, 'welfare may rise when a tax is imposed when entry is taken in consideration' (p. 365). However, he uses a static framework.

© Scottish Economic Society 1996

V Conclusion

This work uses a simple two-period Von Stackelberg model, with sunk costs as a barrier to entry. In such a framework, the potential entrant may decide to enter, even making a loss during the first period, if this loss is offset by second period profits.

As shown, a change in τ modifies the behaviour of the potential entrant, and consequently that of the monopolist. This effect occurs only if $f \in [f^{**}, f^{*}]$ and if there is a positive interest rate. Even by assuming total deduction of sunk costs, profit taxation is not neutral, thus contradicting the result of Davidson and Martin (1991).

This model allows the study of the short-term effects of a profit tax rate change. In terms of an increase in τ, the entrant compares the positive effect due to deduction of losses made during the first period, i.e. $\tau f/(1+r)$, with the negative effect of a greater reduction of resources. If the first effect is greater than the second (i.e. when $q_1^1 > q_1^{1*}$), then an increase in τ is an incentive to enter. In order to prevent the entry, the monopolist must increase production during the first period. Conversely, if the first effect is smaller than the second, then the monopolist reduces production. When τ is reduced, the converse is true. Therefore, the short-term effect of this tax reform on production is ambiguous.

The same ambiguity can be found in the short-term effects on market price and economic welfare. When $q_1^1 > q_1^{1*}$, an increase (decrease) in τ lowers (raises) market price and raises (lowers) economic welfare. When $q_1^1 < q_1^{1*}$, the converse is true. In the long run, an increase (decrease) in τ always lowers (raises) economic welfare.

Acknowledgements

The author is grateful to Gabriella Chiesa, John Ermisch and two anonymous referees for very helpful comments and suggestions. He accepts responsibility for any remaining errors. Special thanks also go to Giacomo Barisone and Louise Finlayson.

Appendix A

In this Appendix the total derivative of net profit with respect to f is computed, under the assumption that $d\tau = 0$. Differentiate net profit $P^1/(1-\tau)$, thus obtaining:

$$d\left(\frac{P^1}{1-\tau}\right) = -\left((2q_1^1 - 1)\,dq_1^1 + \frac{2q_2^1 - 1}{1+r}\,dq_2^1\right). \tag{A1}$$

© Scottish Economic Society 1996

Similarly, differentiate constraint (2):

$$\frac{1}{f^2}\left(\left(\frac{1-q_1^1}{2}\right)^2+\frac{1-\tau}{1+r-\tau}\left(\frac{1-q_2^1}{2}\right)^2\right)$$

$$\times\, df+\frac{1-q_1^1}{2f}\,dq_1^1+\frac{1-\tau}{1+r-\tau}\,\frac{1-q_2^1}{2f}\,dq_2^1=0,$$

which can be simplified and rearranged as follows

$$df=-\left(\frac{1-q_1^1}{2}\,dq_1^1+\frac{1-\tau}{1+r-\tau}\,\frac{1-q_2^1}{2}\,dq_2^1\right). \tag{A2}$$

Furthermore, differentiating equation (8) yields

$$dq_2^1-\mathcal{G}_{q_1}\,dq_1^1-\mathcal{G}_\tau\,d\tau=0, \tag{A3}$$

where

$$\frac{\partial \mathcal{G}}{\partial q_1^1}=\mathcal{G}_{q_1^1}=\frac{\dfrac{(1+r)(1-\tau)}{(1+r-\tau)}}{\left(2(1-q_1^1)+\dfrac{(1+r)(1-\tau)}{(1+r-\tau)}(2q_1^1-1)\right)^2}>0,$$

$$\frac{\partial \mathcal{G}}{\partial \tau}=\mathcal{G}_\tau=-\frac{\dfrac{r(1+r)}{(1+r-\tau)^2}(1-q_1^1)(2q_1^1-1)}{\left(2(1-q_1^1)+\dfrac{(1+r)(1-\tau)}{(1+r-\tau)}(2q_1^1-1)\right)^2}<0.$$

Substituting (A3) into equation (A2) and solving for dq_1^1 yields:

$$dq_1^1=-\frac{2}{1-q_1^1+\dfrac{1-\tau}{1+r-\tau}\mathcal{G}_{q_1^1}(1-q_2^1)}\,df. \tag{A4}$$

Finally, using (A3) and (A4), derivative (A1) can be rearranged as follows:

$$\frac{d}{df}\left(\frac{P^1}{1-\tau}\right)=\frac{2\left((2q_1^1-1)+\dfrac{2q_2^1-1}{1+r}\mathcal{G}_{q_1^1}\right)}{1-q_1^1+\dfrac{1-\tau}{1+r-\tau}\mathcal{G}_{q_1^1}(1-q_2^1)}>0.$$

APPENDIX B

This Appendix finds the sign of $\mathcal{H}_{q_1}-\mathcal{H}_{q_2}$ Recall for simplicity the following

© Scottish Economic Society 1996

partial derivatives

$$\mathcal{H}_{q_1^1} = -\frac{1}{2f}(1-q_1^1),$$

$$\mathcal{H}_{q_2^1} = -\frac{1-\tau}{1+r-\tau}\,\frac{1}{2f}(1-q_2^1).$$

By using them, the above difference becomes

$$\mathcal{H}_{q_1^1} - \mathcal{H}_{q_2^1} = -\frac{1}{2f}\left((1-q_1^1) - \frac{1-\tau}{1+r-\tau}\,(1-q_2^1)\right).$$

Substituting equation (8) into the above equality yields:

$$\mathcal{H}_{q_1^1} - \mathcal{H}_{q_2^1} = -\frac{(1-q_1^1)}{2f}\left(1 - \frac{\dfrac{1-\tau}{1+r-\tau}}{2(1-q_1^1) + \dfrac{(1+r)(1-\tau)}{(1+r-\tau)}\,(2q_1^1-1)}\right)$$

which is always negative, since $\tau \leqslant 1$.

REFERENCES

ATKINSON, A. B. and STIGLITZ, J. E. (1980). *Lectures on Public Economics*. McGraw-Hill, New York.

AUERBACH, A. J. (1989). Tax reform and adjustment costs: the impact on investment and market value. *International Economic Review*, 30, 4, November, pp. 939–962.

BESLEY, T. (1989). Commodity taxation and imperfect competition: a note on the effects of entry. *Journal of Public Economics*, 40, pp. 359–367.

DAVIDSON, C. and MARTIN, L. (1991). Tax incidence in a simple general equilibrium model with collusion and entry. *Journal of Public Economics*, 45, pp. 161–190.

MARSHALL, A. (1890). *Principles of Economics*. Macmillan, London.

MUSGRAVE, R. A. (1959). *The Theory of Public Finance*. McGraw-Hill, New York.

MUSGRAVE, R. A. (1985). A brief history of fiscal doctrine. In A. J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, ch. 1, vol. 1. Elsevier Science Publisher B.V., North Holland, Amsterdam.

MUSGRAVE, R. A. and MUSGRAVE, P. B. (1984). *Public Finance in Theory and Practice*. McGraw-Hill, New York.

PANTEGHINI, P. M. (1995). Monopoly and corporate taxation: the effects of an announced reform. *Rivista Internazionale di Science Economiche e Commerciali*, XLII, 1–2, January–February, pp. 69–91.

STERN, N. (1987). The effects of taxation, price control and government contracts in oligopoly and monopolistic competition. *Journal of Public Economics*, 32, pp. 133–158.

WICKSELL, K. (1896). *Finanztheoretische Untersuchungen und das Steuerwesen Schwedens*. Fisher, Jena.

Date of receipt of final manuscript: 3 May 1995

© Scottish Economic Society 1996

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

THE PRICE–QUALITY RELATIONSHIP UNDER MONOPOLY AND COMPETITION

*Donald A. R. George**

ABSTRACT

The standard economic approaches (screening and signalling) to product quality imply an upward sloping relationship between price and product quality. In the case of screening models it is a relationship which applies to the product line of each individual firm. In the case of signalling models it applies across firms. Both types of model have their shortcomings. An alternative approach is presented which involves endogenising the quality control process. Firms voluntarily provide a warranty contract for their products and this, in turn generates an incentive to control quality. The model compares monopolistic and competitive markets in respect of warranties, quality control and the price–quality relationship. Under monopoly, the price–quality relationship is always upward-sloping or flat. Under competition, however, cost shifts or certain demand side shifts can generate a downward-sloping relationship between price and quality.

I INTRODUCTION

The economics literature on product quality can, broadly speaking, be divided into two categories, screening models and signalling models. Screening models deal with search goods (Nelson, 1970): that is goods whose quality is known to both buyer and seller before purchase takes place. Examples here include the quietness of a car engine or the processing speed of a silicon chip. The firm is assumed to know the distribution of consumers' preferences, but to be ignorant as to the preferences of any given consumer. The firm's problem is thus to screen consumers effectively. Authors who develop screening models include Mussa and Rosen (1978) and Matthews and Moore (1987).

This paper is concerned with a different notion of quality, namely a desirable characteristic of the product which is not known to the buyer before the product is purchased. Examples might be the reliability of the car engine or the durability of the silicon chip. Signalling models do deal with this type of product quality but they are driven by exogenous 'type uncertainty'. That is, the seller does not decide on quality, that is simply dictated by Nature. If product

* Department of Economics, University of Edinburgh

quality were a choice variable for the firm, consumers with full information could simply deduce its value before consuming the good. The signalling literature, deals with a different asymmetric information problem from that of the screening literature. Although the firm knows the quality of its output, the consumer can only discover this quality in the process of consuming the good. Such goods are referred to as 'experience goods'. See for example Grossman (1981), Milgrom and Roberts (1982, 1986), Kreps and Wilson (1982), Klein and Leffler (1981), Shapiro (1983) and McClure and Spector (1991).

Warranties play an important role in both screening and signalling models. Voluntarily provided warranties can help to screen consumers or to signal quality, and in both types of model can serve as a risk-allocating device. It was this last function of warranties which dominated the earlier literature on warranties. For example in Heal (1977) the firm is assumed to produce a probability distribution of qualities. Both firm and consumer know this distribution but neither can tell the quality of an individual product until it is consumed. Heal remarks:

> Typically the quality control (in the production of goods) is sufficiently imperfect that no one will know in advance of their use what their quality will be, and consequently some form of guarantee will be offered.

Heal simply takes the quality distribution as given and does not seek to model the quality control process by which that distribution can be altered. The firm is assumed to be less risk-averse than the consumers and there thus arises a role for warranties as a device by which the firm can insure consumers.

In addition to the economics literature there is an extensive management literature which focuses on techniques of quality control. Issues here include the acquisition and use of information concerning quality. The techniques described are sophisticated and clearly reflect a situation far more complex than is modelled in the economics literature. Real firms evidently do not just select qualities, knowing the quality and production cost of everything they produce. Equally they do not simply produce a probability distribution of qualities devoting no energy at all to influencing quality, and then rely on warranties as a marketing device. The model developed in this paper steers a middle course in assuming that production leads to a probability distribution of qualities, but that firms do devote resources to influencing this distribution. That is, the paper presents a highly simplified model of the quality control process, which reflects the key features of that process which are discussed in the management literature. Whilst this literature presents a vast array of techniques for controlling quality and, to some extent discusses their relative costs and benefits, it does not seek to explain how the firm's quality control decisions are determined. The model presented here treats quality control as part of the firm's overall (expected) profit-maximising decision.

Screening models do not deal with quality of the second type described above and signalling models do not analyse it in a plausible way. The model presented here deals with goods whose quality is unknown to the buyer before purchase, but which is the outcome of a rational decision by the seller. It is a model of

© Scottish Economic Society 1996

imperfect but symmetric information in the sense that both buyer and seller are assumed ignorant of the quality of any particular product, but both know the distribution of qualities in the firm's output.

In order to focus on the quality control process, I abstract from screening (by assuming identical consumers), signalling (by assuming symmetric information), repeat purchases and moral hazard. Section II provides a working definition of 'quality', Section III sets out the model of quality control, Section IV describes the demand side, Section V covers the supply side, Section VI reports results on equilibrium quality control, warranties and pricing under monopoly and competition, and Section VII concludes.

II DEFINITION OF 'QUALITY'

In the management literature product quality is defined extremely broadly. The definition might include price, safety, availability, maintainability, reliability and usability (see e.g., Besterfield, 1986). The focus of interest in this paper is the notion of quality discussed in Section I, namely a desirable characteristic which is unknown to the buyer before purchase. The characteristic chosen is 'durability', which will be defined as: 'the probability of the product not breaking down within the warranty period.' The firm will be assumed to produce goods with different lifetimes. For goods such as light bulbs or silicon chips, which either function or fail to function, this approach presents no difficulties. For goods such as motor vehicles, which are subject to different degrees of breakdown, it is clearly a simplification. The firm will be assumed to produce a range of 'types' of a particular good. A type will be defined by the physical characteristics of the good, but these will not determine the good's lifetime. Goods of a given type will be assumed to have a random lifetime described by a probability distribution. Random factors affecting the lifetimes of light bulbs or silicon chips might include air temperature and mains voltage fluctuations for example. Thus, for a given type of good there is a probability that it will not break down within any given time interval (such as the warranty period). Product quality, in the sense used in this paper, may be thought to be correlated with physical characteristics of the product which are, in principle observable. It will be assumed that the acquisition of precise information on such physical characteristics is too costly for the firm to undertake.

For simplicity 'type' will be modelled as a real number in the interval $[0, 1]$. It can therefore be identified with 'probability of not breaking down within the warranty period'. Throughout the paper the warranty period will, for simplicity, be taken as exogenous, though it might more plausibly be expected to depend, for example, on the stringency of quality control. Also for simplicity the firm will be assumed to produce a uniform distribution of types. Without undertaking any quality control it will be assumed to have no knowledge as to the type of any individual good. The consumer will discover the good's lifetime in the process of consumption but this will not tell her its quality in the sense defined above (i.e. the probability that the good will not break down in the warranty period). Thus the goods in question are neither search nor experience goods in

© Scottish Economic Society 1996

the usual sense of those terms. Note that whether a good is an experience good or not depends on the precise definition of 'quality'. Had quality been defined as 'product lifetime', for example, the product could correctly be described as an experience good.

III Quality Control

The management literature on quality control devotes considerable attention to the techniques available to the firm which enable it to obtain information about the quality of its output. (See for example, Garvin, 1988 and Oakland, 1986.) Such techniques include, for example, product testing, batch sampling etc. It also discusses the uses to which such information can be put. Evidently real production processes generate a distribution of product qualities, and firms apply various techniques to modify this distribution before their products are marketed. The management literature also discusses the way in which product quality information can be used to modify the production process or even the design of products. These issues are suppressed here, in that the produced distribution of qualities is assumed fixed. The unity of the quality control process is often stressed in this literature and there is much discussion of so-called 'total quality management'. In practice however, this often amounts to passing responsibility for quality control back to components suppliers. In any event a rational firm would not produce ultra-high quality products regardless of cost. Quality inspection and reworking of products are therefore unlikely to disappear altogether.

The firm will be assumed to be able to sort its output into products which exceed some quality threshold (the 'stringency level') and those that fall short of it. It will be assumed that the firm can set this threshold at any desired level, and that the direct cost of operating a quality control process depends only on the volume of products it processes. It will further be assumed that goods which fail the quality control process will have some 'rework' value (v) to the firm which is less than the market price of the good (p). For goods (perhaps including silicon chips) which are simply disposed of costlessly when they fail the quality control process, v can be taken to be zero.[1] For goods which require costly disposal v will be negative, while for goods (perhaps including consumer electronics) which can readily be reworked, v will be positive. Thus, a higher stringency level entails a cost to the firm in that the higher the threshold quality it sets the less output it can market for a given level of production and, for each non-marketed product, it incurs a loss of $p - v$. However, if the firm is offering a warranty, higher stringency has the benefit that it will reduce expected warranty costs.

This connection between quality control and warranties has been apparent to managers for some time. Wright (1980) describes events at General Motors:

> I instituted a programme for testing and repairing faulty cars as they came off the assembly line—and the results were phenomenal. It cost about $8 a car,

[1] In fact, 486DX chips which fail the quality control process can often be used as 486SX chips and thus clearly have a positive rework value.

© Scottish Economic Society 1996

> which drove The Fourteenth Floor up the wall. But I figured one way or the other we would end up fixing the defects or paying to have them fixed through recall campaigns or dealer warranty bills ... The internal quality control audit revealed a 66% improvement in the quality of a Chevrolet coming off the assembly line between 1969 and 1973 models. And most important, warranty costs of our new cars were down substantially.

Figure 1 illustrates a quality control process. The firm's total produced output is N and, as mentioned above, it is assumed to produce a uniform distribution of qualities. Total marketed output is clearly given by:

$$\bar{z} = (1 - a)N. \tag{1}$$

Total reworked (or scrapped) output is given by:

$$\bar{y} = aN. \tag{2}$$

The expected quality of a product picked at random from the firm's marketed output is clearly given by:

$$Q = (1 + a)/2. \tag{3}$$

This is, of course, equal to the average probability that any one of the firm's marketed products will not break down during the warranty period.

Although actual quality control processes are extremely complex, the model described in this section does capture their key features in a manner tractable enough for modelling purposes.

IV THE DEMAND SIDE

It will be assumed that there are D risk-averse consumers with identical preferences, each consuming $1/D$ units of the good. (Thus the total marketed output is equal to unity.) In the case of an intermediate good, these consumers

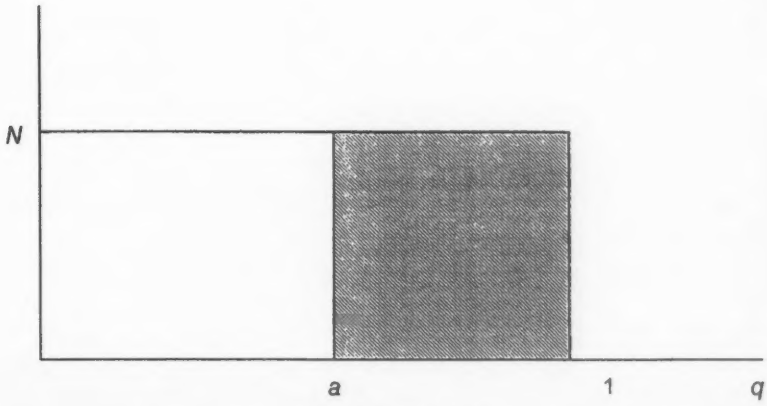


Figure 1. A quality filter.

© Scottish Economic Society 1996

can be thought of as other firms. Buyers know nothing about the quality of the particular goods they are buying, though they can deduce the distribution of qualities across marketed output. All goods are sold under a common standard. The seller offers a contract to the market which consists of a price (p) and a warranty payment (β). This contract requires the seller to make a payment of β to the buyer if the product breaks down within the warranty period. There will be assumed to be no costs of enforcing such a contract. The warranty contract may be thought of as a consumer guarantee or an insurance policy provided jointly with the product or, in the case of an intermediate good, a compensation clause built into the standard inter-firm supply contract. The problem of moral hazard (on the part of buyers) is assumed away.

Each consumer is assumed to have a budget of M and to receive a stream of services from the product. If the product does not break down within the warranty period, this stream of services is worth R to the consumer, while if it does break down the stream of services is worth αR ($\alpha < 1$). Thus the consumer receives income M if she doesn't buy the product, $x = M - p + R$ if she buys and the product doesn't break down, and $y = M - p + \alpha R + \beta$ if she buys and the product does break down. She is assumed to choose a contract so as to maximise expected utility $V(x, y) = \Pi U(x) + (1 - \Pi)U(y)$ where Π is her subjective probability that the product will not break down during the warranty period. Sellers' maximising decisions are clearly subject to the *voluntary participation constraint*:

$$V(x, y) \geq U(M). \tag{4}$$

If this constraint were not satisfied the consumer would clearly not consume the product at all. Note that the consumer is assumed risk-averse, so that $U''(.)$ must be strictly negative. Throughout this paper it will be assumed that the consumers' subjective probability that the product does not break down is equal to the objective probability (i.e. the average quality of marketed output, Q). This average quality is determined by the quality control decisions of sellers and of course the consumer cannot observe these decisions directly. Moreover, the model set up here does not admit repeat purchasing, so the consumer cannot learn about average quality over time. Nonetheless, there is no asymmetry of information so the consumer could in principle deduce the sellers quality control decisions and thus deduce average quality.

V THE SUPPLY SIDE

Quality costs are discussed at some length in the management literature. Groocock (1986) points out:

> Because the products might be defective, they must be inspected and tested. This results in appraisal costs ... Products may also fail a test or inspection, or may fail in the hands of customers. Failure costs are then incurred ... (since the firm) must rework or replace the failed product during manufacturing, or replace or repair the product for customers, for example, under warranty (Groocock, 1986, p. 53).

© Scottish Economic Society 1996

The model developed here formalises these costs as quality control costs, expected rework costs and expected warranty costs. Sellers will be assumed to be risk-neutral maximisers of expected profit. Average and marginal production costs will be assumed constant. Thus the volume of marketed output will be determined by the size of the market (and will thus be normalised to unity, given the assumptions of Section IV above) and the focus of attention is thus placed on the firm's quality control decisions. Moreover, it will be assumed that there are no economies or diseconomies of scale in quality control. Thus the firm's expected profit is given by:

Φ = revenue – production costs – quality control costs
+ rework value of goods which fail quality control
– expected warranty costs.

Note that the rework value might be negative (e.g. for goods which require costly disposal if they fail the quality control process). Recall that N denotes the total produced output of the firm. Then production costs are given by $\gamma_1 N$. Quality control costs are assumed to depend only on quantity processed. In many quality control processes (such as batch sampling) the entire produced output is not processed. Nonetheless, it will be assumed that the quantity processed is proportional to the quantity produced. Together with the assumption of constant (average and marginal) quality control costs, this means that these costs are simply proportional to quantity of output produced. Thus quality control costs = $\gamma_2 N$. It will be assumed throughout that $v < \gamma_1 + \gamma_2$. Expected warranty costs are equal to the warranty payment (β) multiplied by the expected number of products breaking down within the warranty period. Since total marketed output is equal to unity, expected warranty costs must simply equal $\beta(1 - Q)$. Using equation (3), this is equal to $\beta(1 - a)/2$.

Two extreme forms of market structure are considered, monopoly and perfect competition. Quality control decisions turn out to be the same in both cases and the same warranty is offered, but the monopolist charges a higher price. A monopoly equilibrium is a vector (p, β, a) which maximises expected profit (Φ) subject to the voluntary participation constraint (equation (4)). It turns out that this constraint binds in equilibrium. A competitive equilibrium is a vector (p, β, a) which maximises expected utility (V) subject to the constraint $\Phi \geq 0$. This constraint binds in equilibrium (i.e. a zero expected profit condition holds). This may be interpreted as a consequence of free entry or of Bertrand competition.

The level of production (N) and the expected number of marketed products breaking down within the warranty period both depend on the firm's quality control decisions (i.e. on a). Noting that total marketed output is equal to unity, it is clear from equation (1) that:

$$N = 1/(1 - a). \tag{5}$$

Using equation (3), it is clear that expected profit is given by:

$$\begin{aligned}\Phi &= p - (\gamma_1 + \gamma_2)N + vaN - \beta(1 - Q) \\ &= p - (\gamma_1 + \gamma_2)/(1 - a) + va/(1 - a) - \beta(1 - a)/2. \end{aligned} \tag{6}$$

© Scottish Economic Society 1996

VI EQUILIBRIUM AND COMPARATIVE STATICS

Monopoly equilibrium

To characterise a monopoly equilibrium it is necessary to find a vector (p, β, a) which maximises expected profit (equation (6)) subject to the voluntary participation constraint (equation (4)). Taking a multiplier λ for the constraint, this generates the Lagrangian:

$$L(p, \beta, a) = \Phi + \lambda[V(x, y) - U(M)].$$

Noting that $\Pi = Q$, and using equations (3) and (6), this Lagrangian can be written:

$$\begin{aligned} L(p, \beta, a) = {} & p - (\gamma_1 + \gamma_2)/(1 - a) + va/(1 - a) - \beta(1 - a)/2 \\ & + \lambda[((1 + a)/2)U(M - p + R) \\ & + ((1 - a)/2)U(M - p + \alpha R + \beta)]. \end{aligned} \tag{7}$$

Note that $x = M - p + R$ and $y = M - p + \alpha R + \beta$. The first order conditions for an interior solution of the optimising problem require that: $L_p = L_\beta = L_a = 0$. It is now easy to establish:

Proposition 1. In monopoly equilibrium the voluntary participation constraint (equation (4)) is binding.

Proof. Differentiating equation (7) with respect to p gives:

$$L_p = 1 - \lambda[((1 + a)/2)U'(x) + ((1 - a)/2)U'(y)] = 0. \tag{8}$$

Hence $\lambda \neq 0$, and, by complementary slackness the voluntary participation constraint must be binding.

It is now straightforward to prove:

Proposition 2. In monopoly equilibrium the warranty payment (β) is equal to $(1 - \alpha)R$ and thus the consumer receives the same income stream regardless of whether the product breaks down or not (i.e. $x = y$, and the consumer is fully insured).

Proof. Differentiating equation (7) with respect to β gives:

$$\begin{aligned} & L_\beta = -(1 - a)/2 + (\lambda(1 - a)/2)U'(y) = 0 \\ & \Rightarrow \lambda U'(y) = 1. \end{aligned} \tag{9}$$

Using (9) to eliminate λ from (8) gives:

$$U'(x)/U'(y) = 1 \Rightarrow U'(x) = U'(y). \tag{10}$$

$U'(.)$ is invertible since $U''(.)$ is strictly negative, hence equation (10) implies that $x = y$. The result follows.

Proposition 2 simply confirms the standard result that, in equilibrium, a risk-neutral seller will bear all the risk when buyers are risk-averse. It follows from

© Scottish Economic Society 1996

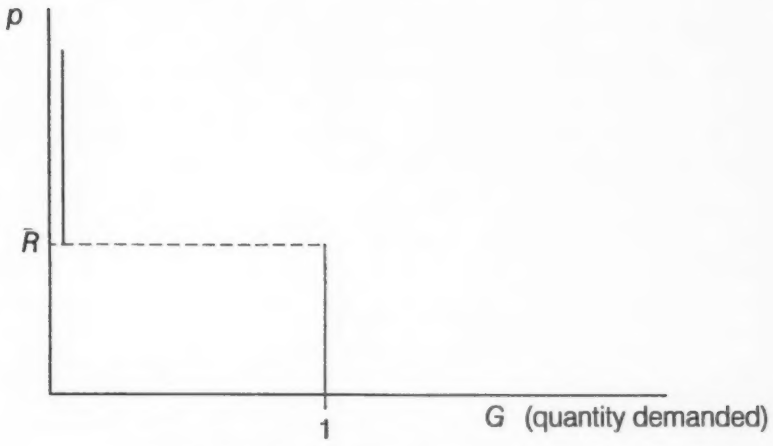


Figure 2. Market demand curve for $\beta = \beta^*$, $a = a^*$.

Proposition 2 that the market demand curve facing the monopolist has the appearance of Figure 2. Combining Propositions 1 and 2, it is easy to establish:

Proposition 3. The monopoly price is given by $p = R$ (i.e. the monopolist extracts all the surplus).

Proof. From Proposition 1 we have:

$$\Pi U(x) + (1 - \Pi)U(y) = U(M).$$

From Proposition 1 we have $x = y$, hence: $U(x) = U(M)$. The function $U(.)$ is invertible because, by assumption, $U'(.)$ is strictly positive. Thus $x = M$. The result follows.

Attention is now turned to the determination of quality control in monopoly equilibrium. Define $\gamma = \gamma_1 + \gamma_2$ and note that $v < \gamma_1 + \gamma_2$ by assumption. We now establish:

Proposition 4. In monopoly equilibrium average quality of marketed output (Q) is given by

$$Q = 1 - \sqrt{(\gamma - v)/2(1 - a)R}. \tag{11}$$

Proof. Differentiating (7) with respect to a and using Propositions 1, 2 and 3, we obtain:

$$L_a = (v - \gamma)/2(1 - a)^2 - \beta/2 = 0$$
$$\Rightarrow a = 1 - \sqrt{(\gamma - v)2/\beta}. \tag{12}$$

Thus, using equation (3) and Proposition 2 the result follows.

© Scottish Economic Society 1996

Competitive equilibrium

To characterise a competitive equilibrium, it is necessary to find a vector (p, β, a) which maximises consumers' expected utility subject to expected profits being non-negative. Taking a multiplier $\mu \geq 0$ for the expected profit constraint, the Lagrangian for this problem is:

$$M(p, \beta, a) = V(x, y) + \mu\Phi.$$

Noting that $\Pi = Q$, and using equations (3) and (6), this Lagrangian can be written:

$$M(p, \beta, a) = ((1+a)/2)U(x) + ((1-a)/2)U(y) + \mu[p - (\gamma_1 + \gamma_2)/(1-a) + va/(1-a) - \beta(1-a)/2] \tag{13}$$

The first order conditions for an interior solution of this optimising problem require that $M_p = M_\beta = M_a = 0$. It is straightforward to establish:

Proposition 5. The expected profit constraint is binding (i.e. $\Phi = 0$) is competitive equilibrium.

Proof. Differentiating (13) with respect to p yields:

$$M_p = -((1+a)/2)U'(x) - ((1-a)/2)U'(y) + \mu = 0. \tag{14}$$

Since $U'(.) > 0$ by assumption, it follows that $\mu \neq 0$. Hence, by complementary slackness, the expected profit constraint binds.

We now prove:

Proposition 6. In competitive equilibrium the warranty payment (β) is equal to $(1-\alpha)R$ and thus the consumer receives the same income stream regardless of whether the product breaks down or not (i.e. $x = y$, and the consumer is fully insured).

Proof. Differentiating equation (12) with respect to β yields:

$$M_\beta = ((1-a)/2)U'(y) - \mu(1-a)/2 = 0$$
$$\Rightarrow \mu = U'(y). \tag{15}$$

Using (15) to eliminate μ from (14) gives $U'(x) = U'(y)$. $U'(.)$ is invertible since $U''(.)$ is strictly negative. Hence equation (14) implies that $x = y$, and the result follows, again confirming a standard result.

We now establish an expression for average quality of marketed output in competitive equilibrium. In fact we have:

Proposition 7. In competitive equilibrium the average quality of marketed output (Q) is given by equation (11).

Proof. Differentiating (13) with respect to a gives (using Proposition 6):

$$M_a = \mu[(v - \gamma)/(1-a)^2 - \beta/2] = 0. \tag{16}$$

© Scottish Economic Society 1996

Noting that $U'(.)>0$, it follows from (14) that $\mu \neq 0$. Hence equation (16) yields the same expression for a as in equation (12) of Proposition 4. The result follows.

The first order conditions for a competitive equilibrium are very similar to those of the monopoly equilibrium. Since the stringency level (a) and the warranty payment (β) are set at the same levels under monopoly as under competition, the market demand curve is the same in both cases. It is depicted in Figure 2. The only difference between monopoly and competitive equilibrium is the product price, which, under competition, is set at a level which forces expected profits to zero (i.e. the non-negative profit constraint is binding). To be more precise, an expression for the competitive price is provided in:

Proposition 8. The competitive equilibrium price is given by:

$$p = v + \sqrt{2(1-a)R(\gamma - v)}. \tag{17}$$

Proof. From Proposition 5 we have $\Phi = 0$ in competitive equilibrium. Note that, by Proposition 7, equation (12) holds for competitive as well as monopoly equilibrium. Thus it may be used to substitute for a in equation (6). Using Proposition 6, β may also be eliminated from equation (6). The result follows.

Comparative statics

From Propositions 2 and 6 it is clear that competition does not induce firms to offer better warranties than a monopolist. From Propositions 4 and 7 it is also clear that competition does not induce firms to undertake more stringent quality control. This is essentially because, under both market structures, *total* utility is maximised, regardless of its distribution. As far as quality is concerned, competition does not provide the consumer with a better outcome than monopoly, but, from Propositions 3 and 8, it is clear that competition does provide that outcome more cheaply.

From Propositions 4 and 7, it can be seen that equation (11) holds for both monopoly and competition, so the comparative statics of quality control are the same for both types of market structure. It is clear from equation (11) that average quality of marketed output is decreasing in the costs (γ_1 and γ_2) of production and of quality control, and decreasing in α. It is increasing in rework value (v) and in the parameter R, which may be thought of as a measure of demand (see Figure 2).

The price–quality relationship is rather more complex however. Both price and quality are endogenous so the observed relationship between them will depend on which parameter is shifting. An upward shift in the demand curve (i.e. an increase in R) will raise price and quality. This is true for both monopoly (Propositions 3 and 4) and competition (Propositions 7 and 8). Thus the price–quality relationship is upward sloping (see Figure 3). An increase in production or quality control costs (γ_1 or γ_2) will have different effects in the two types of market. In both types of market it will lower quality (Propositions 4 and 7) but under monopoly it will have no effect on price (Proposition 3)

© Scottish Economic Society 1996

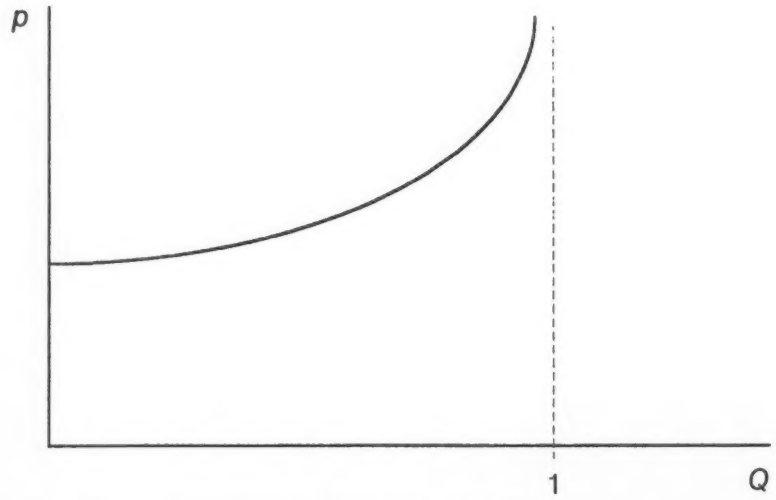


Figure 3. The price–quality relationship induced by shifts in R (monopoly and competition).

while under competition it will raise price (Proposition 8). Thus under monopoly the price–quality relationship is flat (Figure 4) while under competition it is downward sloping (Figure 5). A change in α will induce the same price–quality relationship as a change in γ (Figures 4 and 5). A change in v under monopoly will also induce the price–quality relationship of Figure 4. Under competition however, the price–quality relationship induced by shifts in v cannot be signed.

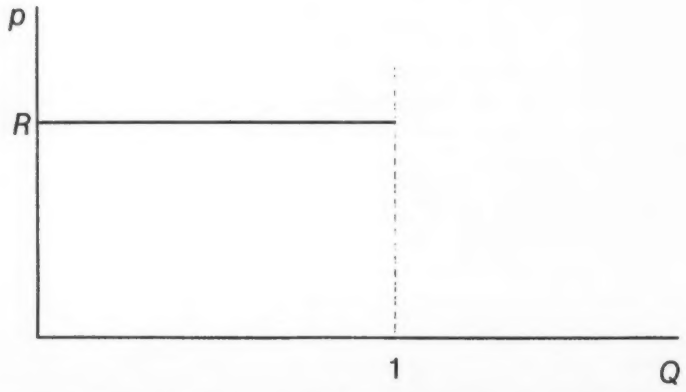


Figure 4. The price–quality relationship induced by shifts in γ, α or v (monopoly).

© Scottish Economic Society 1996

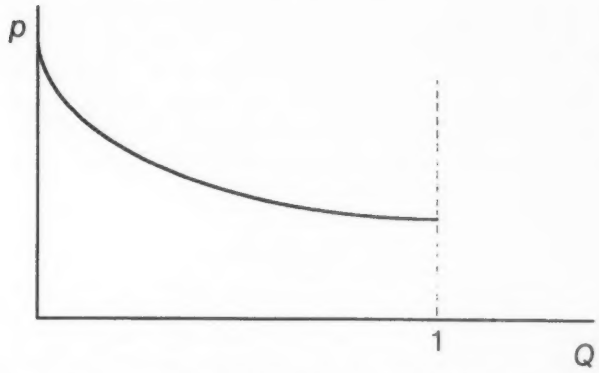


Figure 5. The price–quality relationship induced by shifts in γ or α (competition).

VII CONCLUSIONS

It is frequently the case that product quality is unknown to consumers before they purchase the product. In this situation the determination of product quality and of warranty contracts need not be modelled as the consequence of exogenous type uncertainty. The determination of quality can be treated as an aspect of the firm's (expected) profit maximising decision by explicitly modelling the quality control process. The production process inevitably generates a distribution of qualities and the firm will not know the quality of any individual product. Consumers will not have this information either, so that the goods in question are neither search nor experience goods. Thus the situation analysed in this paper is one of imperfect but symmetric information. This does not mean that firms will market all output regardless of quality. Even though quality control is costly, it will still be undertaken because warranty commitments provide the appropriate incentive. Neither need there be legal compulsion to force firms to offer warranties. A demand for insurance by risk averse buyers will induce risk neutral sellers to offer warranties voluntarily.

A monopolist will offer the same warranty and undertake the same quality control as a competitive firm, but would charge a higher price for the product. Technologies with high rework values of products which fail the quality control process will induce more stringent quality control and thus higher quality of marketed output. Lower production costs or lower quality control costs will have the same effect. Since both price and quality are endogenous in this model, the nature of the price–quality relationship depends on which parameter is shifting and on the type of market structure. Generally speaking the relationship is upward sloping or flat but, in a competitive market, cost shifts and certain demand side shifts can induce a downward-sloping relationship between price and quality.

The analysis of this paper has been comparative static, taking market structure as given. A full dynamic analysis, allowing for reputational effects,

© Scottish Economic Society 1996

might well suggest that a competitive market structure is potentially unstable and could transmute rapidly into monopoly.

The quality control process is complicated and the model presented here inevitably simplifies that process. Nonetheless, quality control is central to the determination of the quality of marketed output, and the economic analysis of quality should therefore include a treatment of this aspect of the firm's behaviour.

ACKNOWLEDGEMENTS

Precursors of this paper have been presented in seminars at Edinburgh University, Queen's University (Canada), Stirling University and the University of East London. The author is grateful for the constructive comments made at those seminars. He is particularly indebted to Simon Clark and John Hardman Moore, and is also grateful for the comments of two anonymous referees. The usual disclaimer applies.

REFERENCES

BESTERFIELD, D. H. (1986). *Quality Control* (2nd ed.). Prentice-Hall.

GARVIN, D. A. (1988). *Managing Quality: The Strategic and Competitive Edge*. Free Press.

GROOCOCK, J. M. (1986). *The Chain of Quality: Market Dominance Through Product Superiority*. John Wiley.

HEAL, G. M. (1977). Guarantees and risk sharing. *Review of Economic Studies*, 44, pp. 549–560.

KLEIN, B. and LEFFLER, K. B. (1981). The role of market forces in assuring contractual performance.*Journal of Political Economy*, 89, pp. 615–641.

KREPS, D. M. and WILSON, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, 27, pp. 253–279.

MATTHEWS, S. and MOORE, J. (1987). Monopoly provision of quality and warranties. *Econometrica*, 55, pp. 441–467.

MCCLURE, J. E. and SPECTOR, L. C. (1991). Joint product signals of quality. *Atlantic Economic Journal*, 19, pp. 38–41.

MILGROM, P. and ROBERTS, J. (1982). Limit pricing and entry under incomplete information: an equilibrium analysis. *Econometrica*, 50, pp. 443–459.

MILGROM, P. and ROBERTS, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94, pp. 796–821.

MUSSA, M. and ROSEN, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18, pp. 301–329.

NELSON, P. (1970). Information and consumer behaviour. *Journal of Political Economy*, 78, pp. 311–329.

OAKLAND, J. S. (1986). *Statistical Process Control*. Heinemann.

SHAPIRO, C. (1983). Premiums for high quality products as returns to reputations. *Quarterly Journal of Economics*, 98, pp. 659–679.

WRIGHT, J. P. (1980). *On a Clear Day You Can See General Motors*. Sidgwick and Jackson.

Date of receipt of final manuscript: 21 July 1995

© Scottish Economic Society 1996

Scottish Journal of Political Economy, Vol. 43, No. 1, February 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

GROSS EXPENDITURE ON UK INDUSTRIAL ASSISTANCE: A RESEARCH NOTE

*Colin Wren**

ABSTRACT

This note presents annual figures for gross expenditure on UK central government assistance to industry over the period 1960–90. This is part of on-going research on the determinants and effects of UK industrial assistance. All figures are at current prices and a breakdown is provided between various components of policy, comprising national assistance, industrial and firm specific assistance, technology support, civil aircraft aid, shipbuilding assistance and local employment and regional policy.

I INTRODUCTION

There exist numerous accounts of the UK experience of industrial intervention (e.g. Denton, 1976; Mottershead, 1978; McCallum, 1979; Grant, 1982; Sawyer, 1991), but to date there exists no single record of the totality of UK expenditure on industrial assistance over a long period. Where figures for spending have been compiled, these usually relate to some component of policy or to some sub-period (for example, Begg *et al.*, 1973, in the case of regional policy). The purpose of this research note is to present annual figures for gross expenditure on UK central government assistance payments to industry, as part of on-going research into the determinants and effects of UK industrial assistance. This series is useful for understanding the scale and composition of industrial assistance at different times, and for examining the growth and decline of these subsidies, and of related phenomena.

We use primary sources to collate figures for gross spending on individual assistance schemes and payments over the period 1960–90. Since grant payments to firms essentially commenced in the late 1950s, and before this amounted to only a few loans at commercial rates (Grove, 1962), this can be viewed as a more or less comprehensive account of central government

* University of Newcastle upon Tyne.

industrial assistance over the post-war period.[1] We take as our definition of industrial assistance all 'unrequited payments to a producer or trader with the objective of reducing the selling price below the factor cost of production' (Feinstein, 1972, p. 80), which not only covers incentives towards the factors of production, but includes lump sum profitability supports used to assist companies in difficulty. As well as grants, we include assistance in the form of loans, interest relief grants, loan guarantees, levies, equity finance and premiums, even though some of this may in principle be recoverable or may have no direct Exchequer implications. Of course, this means that some of the series for assistance expenditure may not be directly comparable with one another, but elsewhere the data are reduced to grant equivalents, enabling a truer comparison of spending across measures and over time (Wren, 1996a).[2] The purpose of this note therefore is to present the data on actual assistance payments.

Related to this last point, we make no effort to determine the value of assistance to the firm, so that the figures relate to gross expenditure.[3] Further, since the value of accelerated depreciation allowances to the firm in the form of initial and investment allowances can differ substantially from that for gross spending, we exclude these as being potentially misleading.[4] We do not include the assistance to agriculture, which has received large-scale support since the Depression, and exclude those activities which may be considered the normal function of government, such as employment schemes involving a significant amount of training or public works, public sector support for defence related projects and that relating directly to trade matters. Finally, our interest is in assistance payments to firms provided by central government, so that we exclude from our analysis those measures made by other levels of

[1] Prior to the war, industrial assistance was generally small-scale. This included assistance under the 1934 Special Areas Act, which amounted to £0.75m in loans and £1.1m in grants, but this was less than that disbursed by the private Nuffield Trust for Special Areas (McCrone, 1969). The 1920s also saw some limited subsidies to the infant civial aviation industry and the establishment of the Unemployment Grants Committee to assist public works schemes, but direct subsidy payments to firms under the latter were rare and usually only made for the purpose of infrastructure provision. Some interventions were made before the First War, including a £2.6 million loan to the Cunard Company for national strategic reasons, and loans and grants to mainly infrastructure projects through the Development Fund under the 1909 Development Act and 1910 Road Improvements Act.

[2] A discussion of the included measures is given in Wren (1996b).

[3] For example, we make no allowance for the fact that prior to 1972 capital grants reduced the capital sum available for depreciation (see McCallum, 1979).

[4] As an example, gross spending on free depreciation on investment in plant and machinery for 1980 is reported at about £10 billion in the *Blue Book*, but allowing for the subsequent reduction in the annual writing down allowance, then under reasonable assumptions actual subsidy in net present value terms is around £3.5 billion (Wren, 1996a). To give some idea of the relative scale of investment tax allowances, spending on these in net present value terms was around £500m per annum over the early 1960s, falling to around £150m in 1970 or about ten per cent of total national assistance (current prices). Following the introduction of free depreciation nationwide in 1972, net present value spending on the tax measures increased steadily to about £3.5bn in 1980, far exceeding that on the measures included here, which sum to about £1.4bn in equivalent terms. The tax allowances were phased out from 1984.

© Scottish Economic Society 1996

© Scottish Economic Society 1996

TABLE 1
Gross industrial assistance expenditure, 1960–1975

(£'m, current prices)

	60/1	61/2	62/3	63/4	64/5	65/6	66/7	67/8	68/9	69/70	70/1	71/2	72/3	73/4	74/5	75/6
National assistance																
Selective employment premium	—	—	—	—	—	—	40·3	129·9	27·4	1·4	0·2	0·2	—	—	—	—
Selective employment tax refund	—	—	—	—	—	—	261·0	706·7	841·9	1253·3	1363·5	998·8	650·8	151·3	1·8	—
Investment grants	—	—	—	—	—	—	—	217·1	340·8	417·1	401·9	401·9	268·7	201·2	103·4	65·4
Industrial support																
Colvilles (R)	15·0	20·0	12·5	2·5	—	—	—	—	—	—	—	—	—	—	—	—
Cotton industry reorganisation	8·4	3·1	3·2	3·3	3·2	2·6	0·1	—	—	—	—	—	—	—	—	—
Industrial Reorganisation Corp. (R)	—	—	—	—	—	—	1·7	15·5	38·3	33·6	30·9	—	—	—	—	—
Aluminium smelters (R)	—	—	—	—	—	—	—	—	18·6	9·4	12·0	9·7	4·4	—	—	—
International Computers Ltd. (R)	—	—	—	—	—	—	—	—	4·4	4·0	3·3	2·3	10·0	10·0	10·0	10·0
Hotels and tourism (R)	—	—	—	—	—	—	—	—	—	4·7	5·7	7·5	11·0	24·4	8·4	2·3
Mineral exploration grants	—	—	—	—	—	—	—	—	—	—	—	—	0·1	0·1	0·2	0·2
Section 8 loans (R)	—	—	—	—	—	—	—	—	—	—	—	—	—	0·8	4·2	4·5
Section 8 grants	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0·8
Section 8 equity (R)	—	—	—	—	—	—	—	—	—	—	—	—	—	5·4	3·0	—
Wool textile industry scheme	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2·2	3·2
Offshore supplies IRG scheme	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0·1	1·1
National Research Development Corp.	—	—	—	2·5	2·7	3·5	7·4	5·5	5·9	6·0	3·5	3·0	2·8	1·4	0·5	0·4
Civil aircraft																
Launching aid (R)	6·9	8·8	13·0	15·9	14·4	3·6	2·7	11·3	9·2	6·9	5·3	0·9	3·7	4·4	15	58
Concorde	—	—	—	—	—	19·2	34·1	44·5	55·7	50·9	64·5	67·1	72·3	68·4	69	68
RB211	—	—	—	—	—	—	—	0·5	10·2	28·2	15·2	69·0	50·1	20·1	45	—

continued

© Scottish Economic Society 1996

TABLE 1
(*continued*)

(£'m, current prices)

	60/1	61/2	62/3	63/4	64/5	65/6	66/7	67/8	68/9	69/70	70/1	71/2	72/3	73/4	74/5	75/6
Shipbuilding assistance																
Shipbuilding credit scheme (R)	—	—	—	—	21·0	30·4	4·8	—	14·1	3·5	—	—	—	—	—	—
Shipbuilders' relief	—	—	—	—	—	—	0·4	3·5	3·6	3·7	4·7	5·2	4·8	6·3	5·8	6·2
Shipbuilding Industry Board loans (R)	—	—	—	—	—	—	—	3·5	5·9	3·6	2·6	3·0	—	—	—	—
Shipbuilding Industry Board grants	—	—	—	—	—	—	—	0·2	5·6	6·5	6·7	2·6	—	—	—	—
Shipbuilding Industry Board equity (R)	—	—	—	—	—	—	—	—	—	3·0	—	—	—	—	—	—
Home credit scheme (R)	—	—	—	—	—	—	—	92·0	198·4	227·2	235·2	91·2	171·3	97·0	226·3	182·2
Other	—	—	—	—	—	1·5	—	—	—	7·2	5·6	—	—	—	—	—
Section 7 loans (R)	—	—	—	—	—	—	—	—	—	—	—	—	6·0	10·6	22·2	29·0
Section 7 grants	—	—	—	—	—	—	—	—	—	—	—	—	1·8	1·1	2·2	5·3
Section 7 equity (R)	—	—	—	—	—	—	—	—	—	—	—	—	5·0	—	19·8	0·5
Construction grant scheme	—	—	—	—	—	—	—	—	—	—	—	—	12·0	18·0	12·0	4·2
Local employment and regional policy																
Building grants	0·1	1·0	1·7	1·4	5·4	9·7	12·6	16·7	17·5	21·5	29·0	30·6	31·3	30·3	15·6	10·9
Discretionary loans (R)	4·6	20·1	12·6	7·5	14·1	9·0	13·4	8·5	17·7	18·4	16·2	14·8	20·4	8·2	0·4	—
Discretionary grants	0·1	0·7	2·5	1·4	1·1	0·3	0·8	0·8	1·2	1·1	0·9	1·2	1·5	0·9	0·6	0·1
Plant/machinery grants	—	—	—	0·1	0·8	2·5	4·4	5·1	1·3	1·6	0·5	0·6	0·1	—	—	—
Investment grants	—	—	—	—	—	—	—	76·9	89·1	98·7	123·1	89·6	67·6	29·5	15·5	5·5
Regional employment premium	—	—	—	—	—	—	—	35·0	103·3	108·8	109·2	109·4	100·9	106·8	155·7	215·1
Selective employment premium	—	—	—	—	—	—	—	—	25·6	27·1	7·4	0·2	0·2	0·3	—	—
Operational grants	—	—	—	—	—	—	—	—	—	—	0·2	2·4	3·8	5·3	4·3	2·5
Regional development grant	—	—	—	—	—	—	—	—	—	—	—	—	8·1	107·2	212·8	324·9
RSA—loans (R)	—	—	—	—	—	—	—	—	—	—	—	—	0·3	29·4	20·1	41·7
—grants	—	—	—	—	—	—	—	—	—	—	—	—	0·1	4·5	15·1	22·3
SIRG/OSIS	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0·2	1·0

© Scottish Economic Society 1996

Sources and Notes:
The table gives gross assistance amounts. (R) indicates that the assistance is at least partly potentially recoverable.
National assistance—Selective employment premium and Selective employment tax refund: *Financial Statistics*, HMSO, London. Investment grants: *Annual Reports on the 1966 Industrial Development Act*. Where higher grant rates apply in the assisted areas, then spending in respect of the higher rate is shown elsewhere.
Industrial support—Colvilles: *Civil Estimates and Estimates for Revenue Departments*, Memorandum by the Financial Secretary to the Treasury, House of Commons 142, Session 1961/62 and *Estimates 1963–64*, Memorandum by the Financial Secretary to the Treasury, Cmnd 1965. The figures for building societies includes £2.2m in payments to housing associations. Cotton industry: *Assistance to the Cotton Industry*, Fourth Report of the Estimates Committee, House of Commons, Session 1961/62 and *Annual Reports*. Industrial Reorganisation Corporation and Mineral Exploration Grants: relevant *Annual Reports*. Aluminium Smelters and ICL: *Annual Reports on the 1968 Industrial Expansion Act*. ICL received £40m in equity over the early 1970s and this has been apportioned to the relevant years. Hotels and tourism: *Annual Reports on the Development of Tourism Act*. Wool textile industry scheme, Offshore supplies interest relief grant scheme and Section 8 payments: *Annual Reports on the 1972 Industry Act*. National Research Development Corporation: *Memorandum by the Financial Secretary to the Treasury*, Supply Estimates, House of Commons, various years. These are estimates from 1965/6 onwards. Shipbuilding and civil aircraft support is shown elsewhere.
Civil aircraft—*Civil Estimates and Estimates for Revenue Departments*, Memorandum by the Financial Secretary to the Treasury, 14th April 1961, House of Commons, Session 1961/62, Mottershead (1978), Table 10.6 and *Government Expenditure Plans*, HMSO, London. Launching aid excludes the amounts made in respect of Concorde and the RB211 project.
Shipbuilding assistance—Shipbuilding credit scheme, Home credit scheme and Other: Mottershead (1978), Table 10.9. The figures for guarantees under the Home credit scheme are based on 80% of the contract value of ships insured under the scheme, but this may overestimate the true position since not all of the guarantees may have been taken up. The Home credit scheme includes guarantees under the 1967 Shipbuilding Industry Act for years up to and including 1971/2. Shipbuilders' relief: *Annual Abstract of Statistics*, HMSO, London. Shipbuilding Industry Board assistance: *Annual Report and Accounts of the Shipbuilding Industry Board*. Section 7 payments and Construction grant scheme: *Annual Reports on the 1972 Industry Act*.
Local employment and regional policy—Selective and regional employment premium: *Financial Statistics*, HMSO, London. Other assistance: *Annual Reports on the 1960 and 1963 Local Employment Acts*, *the 1966 Industrial Development Act* and *the 1972 Industry Act*. The figures for investment grants give grant amounts over and above the standard grant amounts available nationally at 20%, later increased temporarily to 25%. Regional selective assistance (RSA) loans include some equity. SIRG = Service industry removal grant. OSIS = Office and service industry scheme.

© Scottish Economic Society 1996

TABLE 2
Gross industrial assistance expenditure, 1975–90

(£'m, current prices)

	75/6	76/7	77/8	78/9	79/80	80/1	81/2	82/3	83/4	84/5	85/6	86/7	87/8	88/9	89/90	90/1
National assistance																
Investment grants	65·4	26·7	6·1	2·3	1·4	0·5	—	—	—	—	—	—	—	—	—	—
Special employment measures	4	95	181	151	111	377	261	231	118	41	32	27	15	6	—	—
Accelerated projects scheme	—	7·8	14·8	15·3	11·2	3·7	10·9	6·5	0·4	—	—	—	—	—	—	—
Selected investment scheme	—	—	—	3·1	11·0	20·5	32·8	11·9	8·2	5·0	1·9	—	—	—	0·7	—
Support for major projects	—	—	—	—	—	0·2	1·4	4·4	7·2	11·9	10·2	10·3	3·8	2·1	1·4	1·7
Energy conversion	—	—	—	0·1	2·1	6·7	9·3	7·0	—	—	—	—	—	—	—	—
Coal firing scheme	—	—	—	—	—	—	0·1	1·8	4·2	10·1	12·3	11·1	4·3	1·9	1·0	0·1
Consultancy initiatives	—	—	—	—	—	—	—	—	—	—	—	—	—	10·8	35·6	41·2
Industrial support																
ICL (R)	10·0	—	—	—	—	—	200·0	—	—	—	—	—	—	—	—	—
Hotels and tourism (R)	2·3	0·7	0·3	0·1	0·1	—	—	—	—	—	—	—	—	—	—	—
Mineral exploration grants	0·2	0·1	0·1	0·2	0·5	0·7	1·4	0·7	0·5	0·6	0·7	—	—	—	—	—
Section 8 loans (R)	4·5	15·7	14·0	15·0	—	—	—	—	—	—	—	—	—	—	—	—
Section 8 grants	0·8	21·5	2·0	17·5	—	—	—	—	—	—	—	—	—	—	—	—
Sectoral schemes	3·2	7·1	20·4	45·2	50·2	45·8	17·1	6·8	1·6	0·6	0·1	—	—	—	—	—
Offshore supplies IRG scheme	1·1	5·4	11·7	14·3	19·8	27·9	22·2	24·1	22·6	17·7	9·1	4·1	0·3	—	—	—
British Leyland	—	—	325	174	150	300	520	360	90	—	—	—	—	—	—	—
NEB/BTG (R)	—	—	27	39	70	49	39	37	8	—	—	—	—	—	—	—
Private sector steel scheme	—	—	—	—	—	—	—	10·3	14·8	4·4	2·7	—	—	—	—	—
Steel castings rationalisation	—	—	—	—	—	—	—	—	7·1	—	—	—	—	—	—	—
Technology support																
MISP1	—	—	—	—	1·2	1·3	3·7	11·7	10·7	16·4	11·7	0·7	0·2	—	—	—
MISP2	—	—	—	—	—	—	—	—	—	1·8	5·5	7·5	4·6	7·5	2·4	0·8
SEFIS I	—	—	—	—	—	—	—	10·9	8·4	3·8	2·9	—	—	—	—	—
SEFIS II	—	—	—	—	—	—	—	—	15·9	26·3	6·0	—	—	—	—	—

© Scottish Economic Society 1996

Flexible manufacturing scheme	—	—	—	—	—	—	—	—	1·0	4·1	—	—	—	—	—	—
Robot support programme	—	—	—	—	—	—	—	—	1·5	5·1	—	—	—	—	—	—
Advanced manufacturing technology	—	—	—	—	—	—	—	—	—	8·5	10·6	13·7	5·9	4·0	3·1	—
Fibre optics scheme	—	—	—	—	—	—	—	—	8·2	11·6	3·9	4·2	2·6	0·8	0·7	0·3
CADTES	—	—	—	—	—	—	—	—	6·7	11·2	3·9	0·1	—	—	—	—
Innovation linked investment scheme	—	—	—	—	—	—	—	—	—	0·5	0·6	0·3	0·1	—	—	—
Other	—	—	4·2	13·2	29·8	36·4	56·2	65·8	70·0	65·2	109·3	110·7	97·6	80·1	67·4	63·4
Civil aircraft																
Concorde	68	46	48	42	23	36	35	13	—	—	—	—	—	—	—	—
Launch aid (R)	—	—	—	—	—	—	—	10·0	10·7	64·3	81·6	91·7	102·1	99	123	107
Shipbuilding assistance																
Shipbuilders' relief	6·2	8·0	12·8	10·8	11·0	15·8	8·8	18·3	14·0	13·5	22·7	8·4	11·0	8·7	10·3	12·1
Section 7 loans (R)	29·0	0·6	7·7	—	—	—	—	—	—	—	—	—	—	—	—	—
Section 7 grants	5·3	4·0	24·0	2·5	2·4	—	—	—	—	—	—	—	—	—	—	—
Section 7 equity (R)	0·5	0·3	18·0	—	—	—	—	—	—	—	—	—	—	—	—	—
Home credit scheme (R)	182·2	145·9	134·2	205·3	148·1	225·2	298·2	181·7	157·1	130·0	240·1	28·8	77·3	204·4	140·0	213·1
Construction grant scheme	4·2	1·0	0·5	0·3	0·1	0·1	0·3	—	—	0·5	—	—				
Cost escalation scheme (R)	—	13·1	2·3	—	22·6	2·0	0·7	—	—	—	—	—	—	—	—	—
Refund guarantees (R)	—	20·1	12·2	—	—	—	—	—	—	—	—	—	—	—	—	—
Intervention fund	—	1·4	19·3	10·8	32·0	42·0	49·0	48·0	36·2	20·1	25·9	6·6	31·6	51·4	37·7	17·4
Regional policy																
Building grants	10·9	1·5	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Discretionary grants	0·1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Operational grants	2·5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Investment grants	5·5	2·4	0·6	0·6	0·1	0·5	0·2	0·1	—	—	—	—	—	—	—	—
Regional employment premium	215·1	216·8	0·3	—	—	—	—	—	—	—	—	—	—	—	—	—
Regional development grant	324·9	407·7	393·4	416·9	330·8	490·5	616·8	689·6	438·9	410·7	319·7	196·3	118·7	75·5	41·3	18·7
RSA—loans and equity (R)	41·7	21·5	9·1	6·9	1·5	0·1	0·1	0·1	—	—	—	—	—	—	—	—
—grants	22·3	20·6	41·2	91·3	73·7	48·1	75·1	92·1	74·1	78·3	116·1	141·6	154·3	182·2	175·4	188·2
SIRG/OSIS	1·0	1·7	2·2	3·1	3·4	5·4	5·7	5·2	8·8	6·1	6·1	3·3	2·4	1·3	0·1	—

continued

© Scottish Economic Society 1996

TABLE 2
(*continued*)

	75/6	76/7	77/8	78/9	79/80	80/1	81/2	82/3	83/4	84/5	85/6	86/7	87/8	88/9	89/90	90/1
														(£'m, current prices)		
Exchange risk guarantee scheme (R)	—	—	—	42·5	123·6	90·6	62·0	39·1	97·2	72·3	7·5	28·0	54·0	70·0	130·0	101·0
Business improvement scheme	—	—	—	—	—	—	—	—	—	0·2	8·4	14·6	18·4	15·5	11·4	2·1
Revised RDG	—	—	—	—	—	—	—	—	—	—	38·6	116·4	151·9	221·8	162·6	110·9
REG—investment	—	—	—	—	—	—	—	—	—	—	—	—	—	1·1	5·9	8·6
—innovation	—	—	—	—	—	—	—	—	—	—	—	—	—	0·3	2·2	4·0
Consultancy initiatives	—	—	—	—	—	—	—	—	—	—	—	—	—	1·0	3·2	3·7

Sources and Notes:
The table gives gross assistance amounts. (R) indicates that the assistance is at least partly potentially recoverable.
National assistance—Investment grants: *Annual Reports on the 1966 Industrial Development Act*. Special employment measures: Metcalf (1982), Appendix II; *Memorandum submitted by the Department of Employment and the Manpower Services Commission*, Employment Committee, House of Commons 487i, Session 1984/5; *Government Expenditure Plans*, HMSO, London; and *Training Statistics*, Employment Department, HMSO, London, 1990. Included measures comprise the Temporary employment subsidy, the (Temporary) Short-time working compensation scheme, the Adult employment subsidy, the Recruitment subsidy for school-leavers, the Youth employment subsidy, the Young workers scheme and New workers scheme. In the choice of measures we exclude those with a significant training element, and are guided by Lindley (1986), except that the Enterprise allowance scheme is omitted because it was paid to individuals. Other schemes: *Annual Reports on the 1982 Industrial Development Act*. General investment for major projects was superseded by the Assistance for exceptional projects scheme in January 1988. Some of the spending under the Consultancy initiatives at the higher rate has been included in regional spending. Where higher grant rates apply in the Assisted areas, then spending in respect of the higher rate is shown elsewhere.
Industrial support—Hotels and tourism and Mineral exploration grants: relevant *Annual Reports*. ICL, Offshore supplies IRG scheme, sectoral and steel schemes: *Annual Reports on the 1982 Industrial Development Act*. The sectoral schemes include spending on the Wool textile industry scheme and its successor. Assistance to ICL includes a £200m loan guarantee. British Leyland, National Enterprise Board (NEB) and British Technology Group (BTG): *Government Expenditure Plans*, HMSO, London. Shipbuilding, civil aircraft and technology-based support is shown elsewhere.
Technology support—*Annual Review of Government Funded R&D*, HMSO, London and *Annual Reports on the 1982 Industrial Development Act*. Details of payments under some schemes are not available for all years, but the *Annual Review of Government Funded R&D* gives total expenditure in each year, so that the 'Other' heading captures this unattributable spending. Up to 1980/1 this includes the Microelectronics application project and the Product and process development scheme, and later it includes spending under the Alvey programme and other collaborative programmes. MISP = Microelectronics industry support programme. SEFIS = Small engineering firms investment subsidy. CADTES = Computer aided design and test equipment support.
Civil aircraft—*Government Expenditure Plans*, HMSO, London and *Annual Review of Government Funded R&D*, HMSO, London. Launch aid does not include payments made to Rolls-Royce aeroengines in the early 1980s, which at this time was in the public sector.
Shipbuilding assistance—Shipbuilders' relief: *Annual Abstract of Statistics*, HMSO, London and *Annual Report of HM Customs and Excise*, HMSO, London. Other assistance: *Annual Reports on the 1982 Industrial Development Act*.
Regional policy—*Annual Reports on the 1960 and 1963 Local Employment Acts, the 1966 Industrial Development Act, the 1972 Industry Act* and *1982 Industrial Development Act*. Regional employment premium: *Financial Statistics*, HMSO, London. The figures for the Consultancy Initiatives give spending at the higher rate in Assisted areas. Regional selective assistance (RSA) loans include some equity: SIRG = Service industry removal grant. OSIS = Office and service industry scheme. REG = Regional enterprise grant.

government, including the European Union and the activities of development agencies.[5]

II Expenditure on Industrial Assistance

The figures for gross expenditure on industrial assistance are reported in Tables 1 and 2. These value the assistance payments in £ millions at current prices, and provide a breakdown between various components of policy, comprising national assistance, industrial and firm specific assistance, technology support, civil aircraft aid, shipbuilding assistance and local employment and regional policy. These components refer to the availability of aid rather than the intention of policy, since some regional interventions have been made to promote national economic growth, while some sectoral assistance has been made in the context of regional policy. We briefly consider these components in turn.

National assistance

Up to the mid-1970s, Table 1 shows that national assistance comprised three main elements: investment grants, the Selective Employment Tax refund and the Selective Employment Premium. The investment grants were introduced under the 1966 Industrial Development Act and were seen as a more effective way of encouraging investment than the pre-existing tax allowances (Cmnd 2874), leading to the withdrawal of the initial and investment allowances (Dow, 1970). The grants were available on plant and machinery at a rate of 20% nationally, and at a higher rate of 40% in the Development Areas (temporarily raised to 25% and 45% respectively in 1967 and 1968), but were not allowable against the new Corporation Tax, so that the effective rate of subsidy was initially 13% and 26% respectively (Bird and Thirlwall, 1967; Wilson, 1967). The Selective Employment Tax was introduced in the 1966 Budget as a deflationary measure to tackle an emerging trade imbalance and to raise revenue for an expanded public sector. It was levied as a flat-rate surcharge on employers' national insurance contributions and sought a structural adjustment towards those industries with a high export content, so that it was refunded to all firms except those in services and construction. In addition, the Selective Employment Premium was paid in respect of employees in manufacturing and at institutions engaged in related scientific research, but following devaluation it was withdrawn from areas outside the Assisted Areas a year after introduction, and abolished everywhere in 1971.

The investment grants were withdrawn in 1970 in favour of accelerated depreciation allowances (Cmnd 4516), but reappeared in selective form in the

[5] The European Union has provided some direct assistance to UK industry, although much of this has either been used to support infrastructure measures or reimbursed central government for its own spending. The main influence of the EU over the period has been its attempts to limit anti-competitive state aids, although it is likely to play an increased role through the expanded Structural Funds.

© Scottish Economic Society 1996

mid-1970s to tackle a sharp rise in unemployment. The Accelerated Projects Scheme sought to bring large-scale projects forward through time, the Selective Investment Scheme sought to change the nature and scale of investment (Chapman, 1978), and the General Investment Support for Major Projects provided a general facility for assisting projects deemed to be of national importance. The special employment measures from the mid-1970s were a more direct attempt to tackle unemployment, for which details of the included schemes are given in the notes to Table 2. More recently, the Consultancy Initiatives have offered grants to small firms for consultancy-based advisory assistance for purposes designed to raise efficiency.

Industrial support

These were some of the first industrial interventions as economic growth faltered in the late 1950s. Grants for the ailing cotton industry were made available under the 1959 Cotton Industry Act and a £50m loan was agreed for the construction of the Ravenscraig plant by Colvilles in Scotland. Other interventions around this time, such as that for the expansion of the motor vehicle industry, were made under the local employment legislation (Henderson, 1962). The Industrial Reorganisation Corporation (IRC) was formed at the end of 1966 to promote mergers of benefit to the national economy. It was wound up in 1971, and mainly gave loans, but also took share capital. Dissatisfaction with the IRC led the government to take powers to itself through the 1968 Industrial Expansion Act to make loans to establish aluminium smelters at Anglesey and Invergordon, and grants and equity to form International Computers Ltd. through merger (Mottershead, 1978). To improve the balance of payments, loans and grants were made for the construction and improvement of tourist facilities in 1969, and mineral exploration grants were offered from 1972.

Section 8 of the 1972 Industry Act provided a general facility to assist companies in financial difficulty, and was used most notably to assist Chrysler UK, Kearney, Trucker and Marwin and the Meriden motorcycle works. British Leyland was assisted under separate legislation (see Hindley and Richardson, 1983), and this and other governmental holdings, including Rolls-Royce aeroengines, were later transferred to the National Enterprise Board (Parr, 1979). Table 2 shows spending on the various sectoral schemes of the 1974 Labour government, of which the Wool Textile Industry Scheme was an early example. The steel schemes were associated with European Community efforts to rationalise this industry in the early 1980s.

Technology support

In 1977 the Product and Process Development Scheme brought together schemes previously operating under the 1965 Science and Technology Act, offering grant and levy-based assistance towards bringing new processes and products from the design stage to commercial exploitation. As Table 2 shows a

© Scottish Economic Society 1996

number of schemes followed, and this became an important area for policy in the mid-1980s. Much of the activity of the National Enterprise Board was taken up with newly-established technology firms, and in 1981 it was merged with the National Research Development Corporation (NRDC) to form the British Technology Group. The NRDC had channelled government support for research and development since the war, and for a while it offered small grants (Table 1).

Civil aircraft

The civil aircraft industry was a major recipient of government aid in the immediate post-war period, amounting to around £88m by 1958 (Cmnd 2853, Appendix G). This assistance was formalised with the introduction of Launching Aid in 1960, whereby the government agreed to contribute up to 50% of pre-production costs, recoverable by a levy on sales (Hartley, 1974). Both the Concorde airliner and the Rolls-Royce RB211 aeroengine project received funding under this scheme (shown separately in the Tables), but of total Launching Aid of £747m, only £61m was recovered by April 1974 (Gardner, 1976).

Shipbuilding assistance

The shipbuilding industry was an early recipient of state aid to protect it from overseas competition. Initially, cheap loans were offered under the 1964 Shipbuilding Credit Scheme, and indirect tax relief was extended to this sector through Shipbuilders' Relief. Later, reorganisation was recommended (Cmnd 2937), and the Shipbuilding Industry Board was formed to provide assistance with amalgamation. However, many of the new groupings required on-going assistance, including Upper Clyde Shipbuilders, and much of this was provided under section 7 of the 1972 Industry Act (Hogwood, 1979). Part III of this Act introduced guarantees for shipbuilding orders under the Home Credit Scheme, which superseded those previously available under the 1967 Shipbuilding Industry Act, and the Construction Grant Scheme offered a temporary profitability support to mitigate a slump in orders. The Cost Escalation Scheme offered insurance cover in the face of high inflation. To maintain confidence in the face of a protracted nationalisation process, the Refund Guarantees (Performance Bonds) provided government assurances on stage payments and the Intervention Fund offered grants to secure orders. The latter was subject to close scrutiny by the European Commission, which placed increasingly tighter restrictions on this aid.

Local employment and regional policy

Legislation in the immediate post-war period enabled loans and grants to be made to manufacturing firms located or locating in an area of high unemployment, but was little used, so that by 1955 only seventy loans had been made, amounting to £5.9m (House of Commons, 1955). The policy was reactivated in

© Scottish Economic Society 1996

the late 1950s when the conditions on discretionary assistance were weakened, and under the 1960 Local Employment Act a variable rate building grant was introduced (McCrone, 1969). The 1963 Local Employment Act introduced fixed rate automatic grants for the first time, in the form of a 10% plant and machinery grant and a building grant standardised at 25%. Under the 1966 Industrial Development Act a higher rate of investment grant applied to the Development Areas. In response to recession the Regional Employment Premium was introduced in 1967 and was available in respect of those employees eligible for the Selective Employment Premium. The Selective Employment Premium was withdrawn from Development Areas in 1970 to help fund measures in newly designated Intermediate Areas, with increased support given for Special Development Areas through operational grants. The Regional Employment Premium was withdrawn at the end of 1976 (see Armstrong and Taylor, 1993).

Investment grants on plant and machinery were abolished in 1970 and replaced by free depreciation, but reintroduced in 1972 in the form of the Regional Development Grant, along with discretionary assistance in the form of Regional Selective Assistance. At the same time, the free depreciation and a higher rate of initial allowance on industrial buildings were extended throughout the country, and building grants were subsumed within the Regional Development Grant scheme. A revised Regional Development Grant was introduced in 1984 related to job creation, but discontinued in 1988 and replaced by Regional Enterprise Grants and the Consultancy Initiatives. These were based on the Business Improvement Scheme, which was part-funded by the European Union, and offered investment grants and advisory assistance to small firms. From 1978 exchange risk cover was provided on cheap loans under several European Union structural funds towards manufacturing and tourism projects.

III Concluding Remarks

In this note we have given a reasonably comprehensive account of gross industrial assistance provided by UK central government over the post-war period, but excluding expenditure under tax allowances. Inevitably, there may be disagreements over some inclusions and possibly other omissions, but the definition of a producer subsidy must always be contentious (Ricketts, 1985). The presented figures illustrate the different phases of industrial assistance, and show that expenditure on these subsidies in the early 1960s consisted of a few industrial support measures, Launching Aid and limited assistance in localities with high unemployment. The expansion of assistance in the late 1960s occurred in all areas of policy, and notable schemes were the automatic investment grants, the employment premiums and the refund on Selective Employment Tax. In the 1970s, the special employment measures were the main element of nationally available subsidy, while investment grants were confined largely to regional policy. The tables show the gradual withdrawal from industrial assistance in the 1980s, and the final column of Table 2 gives a good

© Scottish Economic Society 1996

indication of remaining measures. Today, industrial assistance has largely returned to the pattern of the early 1960s, consisting of a few industry support measures, limited nationally available assistance, regional aid and support for the civil aircraft industry.

ACKNOWLEDGEMENTS

The author is grateful to two anonymous referees for many constructive comments and to Linda Kelly for valuable assistance, but remains responsible for the paper's content.

REFERENCES

ARMSTRONG, H. and TAYLOR, J. (1993). *Regional Economics and Policy*. Harvester Wheatsheaf, Hemel Hempstead.

BEGG, H. M., LYTHE, C. M., SORLEY, R. and MACDONALD, D. R. (1973). Annual expenditure on special regional assistance to industry, 1960/1–1972/3: a note. *Economic Journal*, 85, pp. 884–7.

BIRD, P. A. and THIRLWALL, A. P. (1967). The incentive to invest in the new development areas. *District Bank Review*, June.

CHAPMAN, J. H. (1978). Department of Industry's accelerated projects scheme and selective investment scheme. *National Westminster Bank Quarterly Review*, May, pp. 27–37.

CMND 2853 (1965). *Report of the Committee of Inquiry into the Aircraft Industry* (Plowden Report). HMSO, London.

CMND 2874 (1966). *Investment Incentives*. Department of Economic Affairs.

CMND 2937 (1966). *Shipbuilding Inquiry Committee* (Geddes Report). HMSO, London.

CMND 4516 (1970). *Industrial Incentives*. Department of Trade and Industry.

DENTON, G. (1976). Financial assistance to British industry. In W. Corden and G. Fels (eds.), *Public Assistance to Industry*. Westview Press, Colorado.

DOW, J. C. R. (1970). *The Management of the British Economy, 1945–60*. Cambridge University Press.

FEINSTEIN, C. H. (1972). *National Income, Expenditure and Output of the United Kingdom, 1855–1965*. Cambridge University Press.

GARDNER, N. K. A. (1976). Economics of launching aid. In A. Whiting (ed.), *The Economics of Industrial Subsidies*. HMSO, London.

GRANT, W. (1982). *The Political Economy of Industrial Policy*. Butterworth, London.

GROVE, J. W. (1962). *Government and Industry in Britain*. Longmans, London.

HARTLEY, K. (1974). *A Market for Aircraft*. Hobart Paper 57. Institute for Economic Affairs, London.

HENDERSON, P. D. (1962). Government and industry. In G. D. N. Worswick and P. H. Ady (eds.), *The British Economy in the Nineteen-Fifties*. Clarendon Press, Oxford.

HINDLEY, B. and RICHARDSON, R. (1983). United Kingdom: pulling the dragon's teeth—the National Enterprise Board. In B. Hindley (ed.), *State Investment Companies in Western Europe*. Macmillan, London.

HOGWOOD, B. W. (1979). *Government and Shipbuilding: The Politics of Industrial Change*. Saxon House, Farnborough.

HOUSE OF COMMONS (1955). *Development Areas*. Select Committee on Estimates, Second Report, Session 1955/6, HMSO, London.

LINDLEY, R. (1986). Labour demand: microeconomic aspects of state intervention. In P. E. Hart (ed.), *Unemployment and Labour Market Policies*. Gower, Aldershot.

MCCALLUM, D. (1979). The development of British regional policy. In D. Maclennan and J. B. Parr (eds.), *Regional Policy: Past Experience and New Directions*. Glasgow Social and Economic Research Studies 6, Martin Robertson, Oxford.

MCCRONE, G. (1969). *Regional Policy in Britain*. George Allen and Unwin, London.

© Scottish Economic Society 1996

METCALF, D. (1982). *Alternatives to Unemployment: Special Employment Measures in Britain*. Report no. 610, Policy Studies Institute, London.
MOTTERSHEAD, P. (1978). Industry policy. In F. T. Blackaby (ed.), *British Economic Policy, 1960–74*, National Institute of Economic and Social Research. Cambridge University Press.
PARR, M. (1979). The National Enterprise Board, *National Westminster Bank Quarterly Review*. February, pp. 51–62.
RICKETTS, M. (1985). The subsidy as a purely normative concept. *Journal of Public Policy*, 5, pp. 401–11.
SAWYER, M. (1991). Industrial policy. In M. Artis and D. Cobham (eds.), *Labour's Economic Policies*. Manchester University Press.
WILSON, T. (1967). Finance for regional industrial development. *Three Banks Review*, September.
WREN, C. (1996a). Grant equivalent expenditure on industrial subsidies in the post-war United Kingdom. *Oxford Bulletin of Economics and Statistics*, forthcoming.
WREN, C. (1996b). *Industrial Subsidies: The UK Experience*. Macmillan, forthcoming.

Date of receipt of final manuscript: 3 April 1995

© Scottish Economic Society 1996

SHORT REVIEWS

S. BAILEY, P. FALCONER and S. MCCHLERY. *Local Government Charges: Policy and Practice*, Harlow, Essex; Longman, 1993, pp. x + 197, £55.00 (hardback).

This book is aimed at managers in government departments and public sector agencies and claims to be 'the definitive guide to the policy and practice of charging'. If any of these managers want definitive conclusions on which aspects of what services should be charged for, and how, they should begin with the last paragraph which concludes that charges 'for public sector outputs will inevitably be fashioned more by political and institutional factors than by economic prescriptions...they must be consistent with service objectives regarding access to, and quality of, service...they must also be capable of being determined within the interactive, sequential and incremental environment of policy-making.' This conclusion is fitting for a book which is more a valuable compendium of useful points about charging than a rigorous 'definitive' study.

Given its objectives, it may be unfair to criticize the book for lacking a careful theoretical statement of the economic arguments about charge finance. But such an overview would have provided a firm basis for the various points made. The opening two chapters do actually give an economic overview, but their overall message seems a little hazy. For instance, we are told that 'there is no legitimate analytical base' for 'the belief that public services are in some sense distinctive and should therefore be provided free at the point of use'; but in case we then assume that they should instead be charged for in the same way as private-sector services, we are told that 'the efficiency rationale for charges is a highly blinkered approach to local government'. In fact, these two chapters suggest that the chief implication of moving from tax finance towards charge finance is a distributional one. This is an interesting argument, so it would have been useful to have some data on the importance of existing or potential charges to household budgets.

One of these economics chapters says that, in the mid-1980s, local authorities levied charges for 600 individual services. In contrast, the following chapter on the politics of charging says that in British public policy as a whole charges are raised in respect of 287 different programmes. The politics chapter also claims that there are greater differences in charging policy and practice between local authorities controlled by the same political party than between councils of different political persuasions; but it seems this interesting observation can be largely explained by the fact that local councils inherit their predecessors' charging policies and cannot make instant reforms. Perhaps the main message of the politics chapter is that charging is a multi-faceted exercise. So the interesting questions are not whether to charge or not, but how charging policies should evolve over time.

After two technical chapters on cost measurement, the book ends with a series of short chapters looking at individual services. These chapters provide interesting summaries of current practices and policy issues. Because the book is a practitioner's manual, these chapters focus on what is currently permissible. So the debate about education vouchers is avoided in the chapter on education charges, even though education is the chief local service, while there is an extended discussion of charging for library services. These chapters will certainly be a mine of information for the intended audience. But the limited theoretical underpinning could be misleading at times. For example, given the importance attached to redistribution, one might hope that a discussion of charges for art galleries and museums would make as much of the fact that charges make people pay in accordance with their usage of the facilities concerned as of the fact that tax-finance benefits typically well-off users. In the discussion of road charging, the idea that congestion charges should be paid into local transport funds deserves more scrutiny; if these charges are intended to bring private motoring costs more into line with social costs, then their destination is irrelevant—except that if they went to local authorities, then those authorities might become more reluctant to ease congestion. And in the discussion of charges for local infrastruc-

ture, we are told that a local property tax would be more closely related than cost-based charges to the benefits from local infrastructure; but aside from the issue of whether people should be charged according to their benefits, this statement ignores the fact that property tax payments depend greatly on the effects of equalization grants.

In short, this is undoubtedly a very useful volume, but it would more fully meet its claims if it had a little more depth.

David King
University of Stirling

V. BROWN, *Adam Smith's Discourse: Canonicity, Commerce and Conscience*, London; Routledge, 1994, pp. 243. £40.00 (hardback), £13.99 (paperback).

The 'Adam Smith Problem'—the alleged inconsistency between the *Wealth of Nations* (WN) and the *Theory of Moral Sentiments* (TMS)—has attracted considerable attention among historians of economic thought. In the present book, V. Brown attempts to give a new interpretation of Adam Smith's texts based on literary theory. In turn the results of the study are used in order to question the received interpretation of Smith's works, the 'canon' from which modern economic analysis traces its descent. The argument is developed in eight chapters. In chapter 1, the author argues that the same discursive context is not appropriate for reading all of Smith's texts. Given this, chapter 2 progresses to utilize M. Bakhtin's work on the development of the novel to deal with the Adam Smith problem. The basic idea here is that TMS may be read as dialogic discourse based on an interplay of moral voices, while the WN may be read as a monologic discource dominated by one didactic voice. Chapters 3 and 4 are concerned with the account of the virtues in the Stoic ethical context which can be found in TMS. Brown argues that in TMS only beneficence and self-command constitute truly moral virtues while justice and prudence are lower order virtues. This ordering of virtues is a manifestation of the influence of the moral hierarchies of Stoicism. Chapter 5 examines justice and juriprudence in TMS; and in *Lectures on Jurisprudence* (LJ) with a view to the concept of the state legislator. The following two chapters examine the shift from TMS and LJ to WN. In WN the mercantile system is shown to be inferior to the system of natural liberty the only coherent and rational economic system in which agriculture should be developed before manufactures and foreign trade. Summarizing the argument, chapter 8 maintains that TMS derogates the virtues associated with the marketplace, and that the WN does not emphasize the benefits of competitive markets and trade, but condemns the overexpansion of trade and manufactures at the expense of agriculture, the more beneficial activity. The author then is able to claim that the TMS and WN are not incompatible, but that they represent Smith's moral contempt for economic activities and therefore calls for a rethink of the received interpretation and of the canon itself.

There have been a number of approaches to the Adam Smith problem. The contribution of this book lies in its attempt to tackle it from the perspective of text analysis. In short, the book is an application of literary theory to the work of Adam Smith. The basic tenet of this approach rests on the underlying assumption that the meaning of a text is not given by the intent of the author but that it is the product of the process of reading. Given the importance of the above idea for the present work, I think that the author should have devoted more space (or even better, a separate section) to discussing the general principles of literary theory. This would have been additionally useful if the book were to be accessible to a wider readership. Generally though, I think that the book offers a refreshing perspective on Adam Smith and, by virtue of its challenge to the mainstream interpretation, a stimulus for further discussion on the subject. It would be valuable to students of economic thought and to all those who wish a new look at the work of the father of modern economics.

S. A. Drakopoulos
University of Aberdeen

© Scottish Economic Society 1996

SCOTTISH JOURNAL OF POLITICAL ECONOMY

Notes for Contributors

The following notes are meant as a guideline for authors when preparing manuscripts for submission to the *Scottish Journal of Political Economy*.

1. **Manuscripts** for consideration should be sent in triplicate to the Editors, *Scottish Journal of Political Economy*, Department of Social and Economic Research, Adam Smith Building, University of Glasgow G12 8RT.
   Authors are requested to type on one side of the paper only, using double spacing with good margins all round on A4 paper.

2. **Footnotes.** A limited number of *explanatory* footnotes are permissible. These should be gathered together, numbered serially and typed on a separate sheet.

3. **References** should be indicated in the text by the surname of the author(s) with the year of publication. References to more than one publication by an author in the same year should be distinguished alphabetically with small letters. The abbreviated author-and-date reference should be placed in parentheses unless the name forms part of the text. The relevant page(s) may be given if necessary.
   Examples:
   (Wolfe, 1973, p. 3); (Little and Mirrlees, 1974a); Laidler (1976) has shown that. . .
   If no person is named as author, the name of the appropriate body should be used, e.g. (Department of Energy, 1975).
   The full reference list should be typed in alphabetical order, double spaced, on a separate sheet of paper at the end of the article. Journal title should be set out in full and if a book is quoted, the publisher, publisher's town and year of publication should be included. Examples:
   Behrend, H. (1957a). The effort bargain. *Industrial and Labour Relations Review*, July 1957, pp. 503–515.
   Balogh, Lord T. (1974). The North Sea Oil blunder. *The Banker*, March, pp. 65–69.
   Lydall, H. (1968). *The Structure of Earnings*. Oxford: Clarendon Press.

4. **Figures and tables** should be submitted on separate sheets and numbered consecutively throughout the article. Figures should, if possible, be drawn on a scale larger than the reproduction required in order to obtain the maximum clarity in reproduction, and submitted for publication either on transparent paper or on white art board. An overlay may be used to indicate the positioning of any words etc. if it is not possible to Letraset directly. Figure legends should be typed on a separate sheet.

5. **Books for Review** should be sent to Professor Alistair Dow, Department of Economics, Glasgow Caledonian University, 70 Cowcaddens Road, Glasgow G4 0BA.

Printed in Great Britain by Page Bros., Norwich.